Nature Reviews Gastroenterology & Hepatology

Artificial Intelligence and Automation in Endoscopy and Surgery

François Chadebecq, Laurence B Lovat, Danail Stoyanov*

* Wellcome / EPSRC Centre for Interventional and Surgical Sciences,
University College London

danail.stoyanov@ucl.ac.uk

Abstract:

Modern endoscopic procedures rely on digital technology ranging from high resolution imaging sensors and displays through to electronics connecting configurable illumination and actuation systems for robotic articulation. In addition to enabling more effective diagnostic and therapeutic interventions, the digitization of the procedural toolset also enables video data capture of the internal human anatomy at unprecedented levels. Interventional video data encapsulates functional and structural information about the patients' anatomy, events, activity as well as action log about surgical process. This detailed but difficult to interpret record from endoscopic procedures can potentially be linked to pre- and post-operative records or patient imaging information.

Rapid advances in Artificial Intelligence (AI), especially in supervised deep learning, can utilize data from endoscopic procedures to develop systems for assisting procedures leading to computer assisted interventions (CAI). CAI systems can provide a wide range of enhanced capabilities, for example better navigation during procedures, automation of image interpretation or even of robotically assisted tool manipulation. In this review, we summarize the state-of-the-art AI for CAI in gastroenterology and surgery.

A. Introduction

Digitization and rapid advances in both hardware and software have been crucial developments that have shaped the capabilities and tools at the disposal of the clinical teams within modern endoscopy suites and operating rooms[1]. These have facilitated the paradigm shift towards minimally invasive surgery (MIS) and procedures that reduce the collateral trauma of interventional care. Camera technologies have been key to allowing visualization of the internal anatomy but without a direct access route, and digital cameras in a range of form factors and configurations are now used across almost all surgical specialties.

The signal captured by cameras/imaging devices provides a rich source of information from the surgical site and it may be captured in a variety of spectrums depending on the clinical specialization ranging from white light images, to narrow band or fluorescence images, all the way through to fluoroscopic or angiographic imaging used in endovascular procedures or to interventional ultrasound. The information within the surgical image or video irrespective of the modality used is akin to a digital record of the intervention and it embeds anatomical information, surgical process and event information, data on instrument use and on the interaction between the instruments and the tissue[2]. With the rapid advances seen in artificial intelligence (AI) over the past decade and specifically in computer vision, it is likely that the next

generation of interventional capabilities will be built upon AI modules that can extract the information from this rich surgical record and provide computer assisted interventions (CAI) in both perioperative and postoperative settings[3-5] (see Fig.1).

While the benefits of MIS over traditional open surgery are well established with shorter hospital stays and recovery times, smaller incision and scars, lower risk of complications and trauma, lower pain and discomfort, and potentially less expensive care; MIS approaches also suffer from inherent practical and clinical limitations:

i. *Perception of the surgical site:* the loss of depth perception, the complex topological and photometric properties of tissue produce blind spots and significant appearance variations and the difficulty to navigate the endoscope make the examination and diagnosis of the gastrointestinal tract difficult. Consequently, important lesions may be missed or misdiagnosed, and tissue areas may be overlooked.
ii. *Endoscope and surgical tool manipulation:* navigating deformable and narrow anatomical cavities is complex and requires a high level of expertise and dexterity in instrument manipulation.
iii. *Analytics and reporting of endoscopic procedures: e*ndoscopy reports are essential to the electronic patient record but limited scope is currently in place for adverse events and any quantitative quality indicators of endoscopic procedures. Detailed clinical reporting is time-consuming and requires standardization (use of comprehensive terminology and evaluation frameworks).
iv. *Multimodality image fusion:* procedures relying on the combined use of different imaging modalities, such as Ultrasound (US) or pre-operative imaging modalities to enable more detailed visualization of the patient anatomy beyond the exposed tissue.

This review intends to cover the emerging trends in technology to alleviate these four critical challenge areas with a predominant focus on AI-based Computer-Assisted Detection and Diagnosis[6-10] (CADe/CADx), CAI and automation.

We define AI as computer methods able to automatically extract information and support decision making by leveraging sensor data and prior knowledge such as labelled datasets for which the expected decision or prediction, called ground truth, is known[4]. Computer vision, aimed at deriving meaningful information from surgical images and videos, is a core component of AI-based CAI platforms dedicated to endoscopy. Most state-of-the art AI methods rely on machine learning, more recently specifically deep learning and artificial neural network which simulate a biological network of interconnected neurons which can be trained to efficiently infer a decision or a prediction given an input signal. Deep learning-based vision systems mirror the human visual cortex and can be designed in a wide range of architectures and node arrangement to be particularly efficient at extracting visual patterns by sequentially applying sets of convolutional filters at different scales[11]. They are the focus of most research in CAI and general medical image computing and analysis[8].

The first section of this review paper will focus on CADe and CADx solutions. Section C will focus on surgical environment mapping and endoscope navigation. Section D will be dedicated to analytics and reporting. Robotically assisted interventions and how AI capabilities can link to novel hardware solutions will be addressed in section E. The regulatory aspects of AI-based medical devices will be discussed in section F as they constitute major impediments to their

operational deployment. The final sections will summarize the review and outline the major challenges to be overcome for effectively deploying AI-based robotic platforms in the OR.

B. Computer-Assisted Detection and Diagnosis

CADe and CADx solutions aim at addressing the challenge of perioperative lesion detection and diagnosis[3]. In endoscopy, they more particularly aim at detecting and classifying areas of abnormal tissue by relying on discriminative visual features. Early CADe and CADx solutions were based on the detection and classification of handcrafted features satisfying prior assumptions on appearance or texture patterns that could be attributed to specific diseases. Conversely, deep learning approaches learn features to discriminate between abnormal and normal tissue based on the data itself without direct assumptions other than the labelled data used to train them, and they classify abnormalities in a much more robust capacity[12].

B.1 CADe: Computer-Assisted Detection

While colonoscopy significantly reduces colorectal cancer mortality, systematic reviews also report that the miss-rate of adenomas could reach 33% for patients at elevated risk of colorectal cancer[13,14]. It has further been established that adenoma miss-rate significantly depends on the quality of bowel preparation, the colonic section explored, the type, size and position of adenomas and the individual practitioner's performance or expertise[15].

A recent systematic review of learning-based polyp detection and segmentation approaches notably reported that AI-based CADe systems significantly alleviate the miss-rate of adenomas in classical white-light endoscopy[16,17]. The most efficient Convolutional Neural Networks (CNNs) achieve an accuracy greater than 95% (proportion of true detection) for the detection of colorectal polyps[18-22]. Another recent systematic review, focusing on the efficiency of modern endoscopy technique for detecting colorectal adenomas, further draws the conclusion that CADe systems overcome new endoscopy imaging technique for the detection of colorectal neoplasia[23].

AI-based approaches have also been successfully applied to upper gastrointestinal endoscopy and more particularly to the detection of neoplasia in Barrett’s esophagus[24]. Recent approaches report a per-image detection sensitivity (true positive rate) and specificity (true negative rate) greater than 90% on the MICCAI (Medical Image Computing and Computer Assisted Intervention conference) EndoVis Challenge Dataset[12,25]. It is notable, however, that the data size was limited and therefore much more work is needed in this application to have confidence in the true performance numbers.

Despite such notable efforts there exist no benchmarks allowing detailed and rigorous comparisons of CADe methods in gastroenterology. The increasing availability of expert annotated datasets allowed the integration of CADe systems within marketed colonoscopy devices (e.g., Olympus ENDOAID endoscopy system, Fujifilm CAD Eye, Medtronic GI Genius, Odin Vision CADU and CADDIE), but additional meta-analyses are required to estimate the impact of CADe on patient care because despite early reports on extremely promising improvements[26] there are many confounding factors that can influence performance including for example human factors and the training time for practitioners to learn to use a system[27].

B.2 CADx: Computer-Assisted Diagnosis

Endoscopy CADx systems aim at classifying neoplastic lesions (see Fig.2). Most CADx approaches rely on enhanced endoscopy imaging capabilities that can interrogate tissue beyond the white light imaging spectrum. The earliest learning-based CADx methods allowed to discriminate between hyperplastic lesions and adenomas relying, notably, on NBI magnified imaging[28-30]. However, despite promising results, they suffered from high computational times which induce latency and impede utility in the clinical setting.

The recent advances in AI and specifically in CNNs have significantly contributed to the improvement of CADx solutions[31]. Recent systems can perform real-time classification of adenomas and hyperplastic polyps in both white light and NBI endoscopy with a classification accuracy reaching 90% for the most effective architectures[32-35]. Relying on NBI has additionally shown that CNN-based CADx approaches can distinguish between 5 different classes of colorectal lesions[36,37] (MS classification, MSI, MSII, MSII0, MSIIIa, MSIIIb). Similar approaches have successfully been applied to endocytoscopy with comparable classification accuracy results[38,39]. The joint integration of CADx and CADe systems within CAI platform has also recently become possible with some emerging commercial systems providing this capability (Odin Vision CADDIE).

While state-of-the-art CADx systems' performance compares to human experts for colorectal cancer diagnosis, the lack of large-scale datasets that have been annotated by clinicians and benchmarks tailored to the different neoplastic classification standards used in gastroenterology, make the clinical evaluation of these systems difficult[40-42] (e.g., robustness, limitations of CADx system in white light endoscopy, ability to classify neoplastic lesions according to the Paris classification). The CADx system Endo-Brain, initially developed by Cybernet System corporation, was the first system to receive a regulatory approval (approbation delivered by the Japan’s Pharmaceuticals and Medical Devices Agency, see section F).

C. Endoscopic Mapping and Navigation, Anatomical Structures Identification, and Instrument Segmentation

The 3D mapping of the endoscopic scene and the localization of the endoscope within this environment are essential to CAI and support for navigation in endoscopy. Combined with the semantic analysis of endoscopic scenes, such CAI platforms could improve endoscopic procedure safety, allow the development of endoscopy reporting systems[43] and even play a critical role in the development of multi-modal endoscopy.

C.1 Endoscopic Mapping and Navigation

Most Simultaneous Localization And Mapping (SLAM) approaches dedicated to endoscopy rely on complex photogrammetry pipelines aiming at simultaneously inferring the geometry of the endoscopic scene and the endoscope’s displacements directly from a sequence of images[44] (see Fig.3).

Conventional SLAM methods assume the scene to be rigid which limits their applicability in deformable endoscopic applications. Additionally, detection and tracking of visual features can be challenging in endoscopy and MIS where there can be a paucity of salient tissue texture and appearance can vary with clinical events like bleeding (See section A, i). ORB-Slam, and its variants dedicated to endoscopy, remain the current gold standard SLAM approaches in white light endoscopy[45-49]. Such systems mainly differ in image matching strategies and recent approaches can achieve real-time dense reconstruction of endoscopic scenes[48,49]. Despite the

development of advanced deformable SLAM pipelines[50-53], the reliable tracking of visual features remains challenging and prevents the implementation of conventional SLAM approaches in robust and stable clinical systems. Active vision techniques, notably relying on the use of electro-magnetic tracking devices or structured illumination systems[54], only partially compensate for visual endoscopy artefacts and the fusion of heterogeneous sources of data remains difficult.

CNN-based SLAM pipelines avoid the need for tracking visual features by directly estimating the depth map of the endoscopic scene from a single monocular view (see Fig.3). State-of-the-art CNN-based pipelines[55,56] favorably compared to[45,48], notably by reporting accurate mapping and localization results on long duration colonoscopy sequences. They however remain particularly sensitive to endoscopic imaging artefacts and their robustness to deformation is limited[56]. Despite noteworthy progress, the lack of large-scale annotated datasets, especially using simulation environments[57,58], is a major bottleneck to the advance of learning-based visual SLAM systems and arguable one of the main reasons why currently there is no reliable solution to navigation in gastrointestinal endoscopy.

C.2 Anatomical Structure Recognition

Anatomical structure recognition may consist in detecting different sections of the gastrointestinal tract but also detecting critical structures or landmarks within these sections. While similar to CADx and CADe system, in this section we consider structures and methods that have not focused on polyps in colonoscopy.

Different classifications of the upper gastrointestinal tract are utilized in clinical practice[59,60]. Derived from the British and modified Japanese guideline[61-63], a new dataset distinguishing between 11 locations of the upper gastrointestinal tract (pharynx, esophagus, squamocolumnar junction, middle-upper body of antegrade view, lower body, antrum, duodenal bulb, duodenal descending, fundus, middle-upper body of retroflex view, angulus and a 12^{th} class associated to unqualified landmark) has been proposed[64]. This study further demonstrated that conventional CNNs architectures perform equally well with an average classification accuracy greater than 85%. Considering a relatively similar classification, but distinguishing between white light and NBI esophagogastroduodenoscopy, a classification accuracy greater than 95% has been reported using another conventional CNN architecture[60].

Similar studies have been proposed for anatomical site segmentation in colonoscopy. A longitudinal analysis of the performances of reference CNN approaches[65] on three reference colonoscopy image classification challenges[66-68] has recently been performed. The most recent challenges[68-70] aim at evaluating image classification methods for distinguishing between anatomical location (z-line, pylorus, cecum, retroflex rectum, retroflex stomach), abnormalities (polyps, ulcerative colitis), polyp removal cases (dyed and lifted polyps, dyed resection margins), and surgical context (normal colon mucosa, moderate stool inclusion, significant stool inclusion, useless blurry image, surgical instrument detected, out of patient). Most state-of-the-art CNN-based approaches[71] reach an accuracy greater than 90%.

Despite promising results, the reliability and accuracy of CNN-based anatomical structure recognition methods highly depend on the chosen classification criteria as well as its granularity and the datasets used to train the models. More particularly, identifying different sections of the colon; the terminal ileum, the cecum, the ascending to transverse colon section, the descending

to sigmoid colon section, the rectum, the anus (and a class associated to indistinguishable parts of the colon) remains challenging[72]. Applications such as automatic endoscopy reporting require a consistent level of description to be usable in clinical applications.

C.3 Surgical Instrument Detection

Most research focused on surgical tool segmentation has been in laparoscopy rather than endoscopy and especially in robotically assisted laparoscopy making progress towards robot autonomy. Several open datasets for training AI models have allowed state-of-the-art methods to rely on supervised CNN-based semantic segmentation methods[73], though notable recent progress has also been made using image synthesis and image-translation. The ROBUST-MIS 2019 challenge[74] highlighted the efficiency of such methods for segmenting surgical tools but also their lack of robustness in segmenting small tools orientated in the view axis of the camera, rapidly moving, overlapping, or crossing instruments[74-76]. Recent research on the development of unsupervised approaches for the segmentation of surgical tools has also been promising showing that synthetic images can successfully be used to train CNNs for this application[77-81]. It has further been shown that meta-learning methods can allow their adaptation to diverse types of surgical instruments[82].

Surgical tools pose estimation requires segmenting their multiple articulations which remains a challenging task[74]. Several CNN-based approaches aim at jointly estimating the segmentation and 2D pose of surgical tools[83,84]. Despite promising results, notably to predict occluded instruments joints, further studies need to be carried out to evaluate their efficiency in real endoscopy scenario. Other approaches rely on generalizable models of articulated surgical instruments to infer their 3D pose in robotic-assisted surgery[85]. By efficiently combining a CNN jointly detecting, segmenting, and extracting landmark primitives of multi-articulated surgical tools with a geometric method allowing the estimation of their 3D poses, it has been shown[86] that the tip of surgical tools can be located with a mean error of 3 mm albeit in limited circumstances.

Despite significant advances, multi-part surgical tool segmentation remains challenging and the lack of reliability of SLAM approaches limits its relevance.

D. AI for Understanding Surgical Process

The identification of surgical workflow or activity is key to automated endoscopy reporting and it could play a significant role in the automation of surgical procedures (e.g., by notably allowing the generation of large annotated datasets). Further such process understanding can be associated with surgical skill assessment systems and, it may contribute to the development of CAI platforms for training in endoscopy[87].

D.1 Surgical Workflow Recognition

Most research in recognition of procedural understanding has focused on laparoscopic surgical workflow analysis, predominantly through automatic decomposition of surgical procedures into actions at different granularity levels such as gestures, activities, or phases (from a low level to a high-level decomposition)[88-91]. State-of-the-art approaches are based on architectures that incorporate temporal information with models that can be solely based on image/video data or integrate multi-modal data (e.g., sensor information such as robot kinematics).

Fine grained gesture recognition approaches so far have been heavily reliant on the JIGSAWS dataset[87] despite it being limited to ex-vivo procedures mimicking robotic surgery suturing tasks

in silicone training phantoms (see Fig.4). Examples of gestures are “pushing a needle through the tissue” or “transferring a needle from left to right”. A systematic review highlighted the prevalence of supervised CNNs-based approaches[88]. Nevertheless, temporal NNs integrating multi-modal data[91-96] (e.g., video, robot kinematic, surgical tools identification) achieved a per-image surgical gesture recognition accuracy of approximately 90%. A recent approach notably embedded multimodal attention mechanisms within a two-stream temporal network to efficiently combine kinematic and video data[97]. It achieved higher accuracy and better consistency than unimodal solutions on both phantom and in-vivo data.

Activity recognition approaches at a coarser level mainly rely on the availability of laparoscopic video from the TUM LapChole, cholec80 and M2CAI datasets which focus on laparoscopic cholecystectomy[98,99]. Examples of activities include “calot triangle dissection” or “clipping and cutting”. Unsurprisingly, temporal NNs significantly outperform conventional CNNs methods by achieving a recognition accuracy of approximately 90% on the cholec80 dataset[100,101]. Analysis of the MIDL (Medical Imaging and Deep Learning conference) 2020 SARAS-ESAD challenge found similar conclusions for activity recognition in robotic-assisted prostatectomy[102,103] albeit on a fairly small dataset. Laparoscopic colorectal surgery activities recognition has also been investigated[104] with examples of activities such as “lateral mobilization of colon” (approach to mesocolon from lateral side) or “TME (left side)” (approach to mesorectum on the left side for dissection). Preliminary analysis using a state-of-the-art CNN-based approach demonstrates that a recognition precision of approximately 80% can be achieved. Beyond action recognition, action prediction has also been explored by a recent approach, based on a temporal NN, achieving a laparoscopic cholecystectomy action prediction accuracy of approximately 60%[105]. With the aim of providing comprehensive fine-grained information on laparoscopic cholecystectomy surgical activities, the CholecT50 dataset[106,107] (mainly relying on cholec80 dataset videos) is annotated with triplet information in the form of <instrument, verb, target> (i.e., an instrument will be used to perform a specific action on a target organ). The analysis of the 19 state-of-the-art approaches competing at the endoscopic vision challenge organized at MICCAI 2021 shows that surgical workflow analysis remains unsolved [107] (with a mean average precision only ranging from 4.2% to 38.1%).

The definition of surgical phases and activities is challenging and remains ambiguous and at times as shown in laparoscopic cholecystectomy, these terms can be confused[97,101]. The lack of standardized definition of phases is an impediment to the development and evaluation of gastrointestinal endoscopy phase recognition approaches but also perhaps reflects that this need did not exist previously. To overcome this issue, recent studies focus on the definition of standard ontologies for endoscopic procedures[91,108,109]. More particularly, the Heidelberg colorectal dataset (ROBUST-MIS Challenge[74]) can be used to identify standard phases in three different laparoscopic procedures[91] (proctocolectomy, rectal resection, sigmoid resection).

The lack of standardized definition of gestures, activities, and phases of gastrointestinal endoscopy procedures as well as the difficulties to generalize surgical workflow approaches to different procedures at different granularity levels remain critical open problems. If contextual information such as the practitioners’ position within the operating room or the analysis of practitioners’ comments could benefit surgical workflow analysis, the integration of multi-modal information remains challenging[7,110].

D.2 Surgical Skill Assessment

Surgical skill assessment efforts have mainly focused on the automatic evaluation of practitioners' expertise in performing specific surgical tasks. Being closely related to surgical workflow analysis, the automation of surgical skill assessment has significantly advanced with the emergence of temporal NNs. Such networks have notably been trained to distinguish between three levels of practitioners' expertise (novice, intermediate and expert) on the three tasks of the JIGSAWS dataset[111] and, in a more elaborate way, to jointly recognize surgical gestures and evaluate skill scores[112] (ranging from 6 to 30). Both methods achieve skill assessment accuracy greater than 95% but the lack of real surgical data evaluation is a major limitation to the validity and impact of such efforts[113]. A three-stage temporal NN-based approach dedicated to laparoscopic cholecystectomy has also been proposed[114]. It achieves, on a private cholecystectomy dataset, an average classification accuracy of approximately 85% in distinguishing good and poor surgical skills. More recently, a unified learning-based approach has been presented to exploit and combine distinct aspects of surgical skill such as the identification of surgical instrument usage or intraoperative event patterns[115]. The proposed method outperforms the state-of-the-art for estimating skill scores (ranging from 7 to 35) on in-vivo laparoscopic gastrectomy and lymph node dissection. Nevertheless, large-scale clinical datasets are required to provide evidence of the reliability of this approach.

Beyond the evaluation of practitioners' expertise, surgical skill assessment methods also focused on the derivation of quality metrics in endoscopy[116]. CNNs have notably been trained to automatically evaluate pre-operative bowel preparation quality[117], which is important as a means of assessing the efficacy of the procedure. Other prospective studies focused on the detection of erroneous surgical gestures, such as erroneous motions of surgical instruments, as they could significantly contribute to surgical trainees' performance improvement[118].

Despite promising results, the automation of surgical skill assessment remains limited by the lack of large-scale expert annotated datasets and suffers from the same bottleneck as surgical workflow recognition. Recent efforts in releasing new data, particularly in simulation settings, are an important activity that will benefit the community[119].

E. Robotic Assistants and Automation

The robotic system STAR, "Smart Tissue Autonomous Robot", was the first robotic system able to perform a suturing procedure with a minimum amount of guidance[120]. It has thereafter been significantly improved by integrating specialized suturing tools and advanced imaging systems beyond the visible spectrum. Recently, the STAR system has demonstrated feasibility performance (on both phantom and animal models) for laparoscopic suturing of bowel anastomosis, a complex soft-tissue surgical task requiring a high level of both accuracy and consistency to prevent the risk of anastomotic leakage[121].

Despite the growing adoption of minimally-invasive robotically-assisted surgery, it has been reported that surgeries performed using the Intuitive surgical system, which is by far the global market leader, represent less than 10% of the overall soft tissue surgery in the USA and less than 0.5% of all surgery globally[122,123]. Major impediments to the generalization of robotically assisted surgery are the costs, the practitioners' learning curve, the operative time, the limited number of surgical tools and imaging modes, the lack of evidence that robotically assisted surgery outperforms conventional and MIS for numerous procedures.

As presented in the previous sections, AI-based CAI platforms could be a key element in overcoming some of the current impediments to more effective and intelligent robotic surgery[124-

[126] (see Fig.5). Robot feedback systems and kinematic data from the encoders could mitigate the limitation of vision-based AI such as computer-assisted navigation platforms by providing an additional sensor. Combining sensor data stemming from different sources remains, however, a difficult and open challenge. State-of-the-art approaches utilizing data fusion notably investigate surgical subtasks automation, such as suturing[127], tissue cutting[128] or tissue flaps retraction[129]. Image-guided endoscope navigation is also the object of much research[130] (see also section C.1). An increasing number of studies further address the challenge of autonomous soft tissue manipulation by predicting soft tissue deformations to enable planification of surgical instrument motion[131]. Preliminary evaluations of these methods, carried out on synthetic data and animal models, demonstrate promising results but further investigations are needed to provide evidence of their robustness. Even though these methods were applied to laparoscopy, most of them could be adapted to flexible and robotically-assisted endoscopy[132].

Beyond the automation of conventional endoscopy, AI could also contribute to the development of new robotic platforms and paradigms. Magnetic endoscopy is notably presented as a promising alternative to colonoscopy although human-machine interfaces are needed for providing practitioners with spatial and contextual information. Prospective studies aim at integrating robot feedback systems and computer vision to design magnetic endoscopy platforms[133]. Evaluation of such systems carried out on animal models demonstrates the feasibility of the proposed approach and opens new perspectives towards the development of autonomous colonoscopy systems.

The lack of appropriate regulatory frameworks and evaluation standards prevent a thorough assessment of the resilience, robustness, accuracy, and reliability of AI-driven surgical robotic platforms[125]. They are essential to demonstrate the effective benefit of autonomous minimally invasive surgery over minimally invasive surgery[134-136]. Further work is also needed to understand the clinical value and healthcare economics of such autonomous capabilities.

F. Regulatory Approval and Reimbursement Schemes

Beyond technical restrictions, the operational implementation of AI-based CAI platforms requires regulatory and ethical guidance, legal responsibility frameworks and appropriate reimbursement schemes. Regulatory approval pathways are globally evolving but countries differ in their approach to regulation[136,137]. As of 2022, 13 AI devices have cleared regulatory approvals in Europe, China and Japan[138,139]. These AI devices are mainly dedicated to the detection of polyps in colonoscopy (e.g., Olympus ENDO-AID, Odin Vision CADDIE). The approbation of AI-based platforms notably depends on their relative impact. A decisional system that could cause irreversible or serious deterioration of a patient's state of health or a surgical intervention will fall in the higher class of risks for software as a medical device (SaMD). The level of certification associated with each class of risk depends on local regulatory organizations. CADe platforms fall in class II in Europe and in the USA (over 3 classes) but, a third-party certification is required in Europe (clinical validation based on scientific validity, analytical validation and clinical validation and a CE certification to access the access the European market) while an FDA approval is required in the USA (mainly based on estimates for detection through a premarket Notification 510(k) process[140]). The CADx system Endo-Brain has been approved by the Japan's Pharmaceuticals and Medical Devices Agency as a Class III device[141]. As highlighted in previous surveys[136], even though regulation policies are evolving, the uncertainty around requirements is a major impediment to the implementation of AI devices in gastroenterology.

Together with the adaptation of regulatory approval procedures for AI-based medical devices, legal frameworks must be defined to address critical liability issues[139] and establish appropriate reimbursement schemes. Currently, the lack of evidence regarding AI-based devices cost-saving prevents any refund charges by public health organizations or health insurance systems[142,143]. Despite recent studies demonstrating the cost-efficiency of AI-based devices in MIS, additional evidence is required to reach a clinical consensus[142]. Consequently, AI-based devices are currently used as assistive systems and the implementation of autonomous decisional or interventional system remains hypothetical.

G. Discussion, Challenges and Focus of Development

If AI-based CAI platforms could enhance precision, link to robotic instrumentation, and allow augmented visualization of the surgical site, a range of clinical and technical bottlenecks yet remain to be overcome for CAI platforms to demonstrate impact on improving patient care. Some of the major technical challenges are:

i. Most practical AI-based methods for CAI rely on the supervised training of CNN models using large-scale annotated datasets. Despite the increasing availability of endoscopic video, labelling physiologically meaningful structures requires expertise. Also, for some perception problems such as surgical environment navigation, it is difficult, if not impossible, to collect labelled data (e.g., 3D spatial labels).
ii. Unsupervised training of CNN models is a challenging computer vision problem. AI-based CAI approaches mainly rely on weakly-supervised learning strategies. Synthetic data and transfer learning, allowing a network train on a source domain to adapt to a target domain, are independently or jointly used. It remains, however, difficult to generalize learning-based approaches to various endoscopic scenarios.
iii. Generalizing AI methods and designing scalable CAI solutions able to adapt to the peculiar anatomical and physiological characteristics of a patient, the broad range of anatomical abnormalities appearances and physiological manifestations of diseases, as well as different endoscope imaging system properties is complex. This problem implicitly extends to the ability of a CAI platform to adapt to singular environment and scenarios seldom or never represented within the training dataset.
iv. Efficiently combining heterogeneous multimodal data remains an open computer vision problem. Heterogeneous data such as endoscopic images and kinematic data stemming from a robotic platform lie in different domains. Meanwhile, multimodal images, collected pre and per-operatively, display significantly different and seldom overlapping features.

Clinical challenges for the deployment of AI-based CAI solutions can be summarized as:

v. While CNN inference is fast, the inference of deep neural network models does not always satisfy real-time constraints. Efficiently integrating multiple functionalities within a CAI platform, and more particularly efficiently combining learning-based and physics-based models, remains a critical problem that is often overlooked. It imposes a trade-off between the tractability and the robustness and accuracy of the proposed CAI solution. Promising efforts, for example by NVIDIA and the Clara AGX, could provide platform technologies for addressing the computational needs for systems[144].
vi. Various levels of regulation are needed for integrating medical devices and software within the OR. AI-based solutions rely on complex models that are difficult to interpret. As such, assessing the clinical limitations and capabilities of such models is difficult,

particularly for problems in which human supervision cannot be used to validate their precision such as surgical navigation. Additionally, regulators may ask for data to be representative of different population demographics, which needs to be considered from the early stage of the AI model development.

vii. For deploying AI-based CAI platforms within the OR, it is critical to provide the surgical team with simple human-machine interface that will provide them with clinically relevant information while allowing them to effectively communicate with these platforms. This notably involves integrating contextual information such as the recognition of surgical phase being performed. The robustness of system performance needs to be valid across heterogeneous aspects of the clinical environment, such as the surgical approach, patient specific information, different instrumentation or devices that might be utilized.

In this review paper, we have tried to succinctly summarize current progress in AI systems for endoscopic video analysis and some perspectives on how these may impact endoscopic robotics. The constant advances of AI and the increasing availability of expert annotated datasets notably allowed the recent integration of AI CADe and CADx devices in gastroenterology. Nevertheless, critical technical and clinical challenges significantly hinder the implementation of autonomous decisional and interventional AI-devices. Despite promising perspectives, intrinsic limitations of current learning-based approaches will need to be overcome for AI-devices to become key components of modern surgical capabilities.

References

1. Darzi, A. & Munz, Y. The Impact of Minimally Invasive Surgical Techniques. Annual Review of Medicine. 55, 223-237 (2004).

2. Stoyanov, D. Surgical vision. Annal of Biomedical Engineering. 40, 332-345 (2012).

3. Ahmad, O.F. et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. The Lancet Gastroenterology & Hepatology. 4, 71-80 (2019).

4. Kaul, v., Enslin, S. & Gross, S.A. History of artificial intelligence in medicine. Gastrointestinal Endoscopy. 92, 807-812 (2020).

5. Jin, Z. et al. Deep learning for gastroscopic images: computer-aided techniques for clinicians. BioMedical Engineering OnLine. 21 (2022).

6. Maier-Hein, L. et al. Surgical Data Science-from Concepts to Clinical Translation. Medical Image Analysis. 76 (2022).

7. Maier-Hein, L. et al. Surgical data science for next-generation interventions. Nature Biomedical Engineering. 1, 691-696, (2017).

8. Vercauteren, T., Unberath, M., Padoy & N., Navab, N. CAI4CAI: The Rise of Contextual Artificial Intelligence in Computer-Assisted Interventions. Proceedings of the IEEE. 108, 198-214 (2020).

9. Le Berre, C. et al. Application of Artificial Intelligence to Gastroenterology and Hepatology. Gastroenterology Journal. 158, 76-94 (2019).

10. Chadebecq, F., Vasconcelos, F., Mazomenos, E. & Stoyanov, D. Computer Vision in the Surgical Operating Room. Visceral Medicine. 36, 456–462 (2020).

11. Goodfellow, I.J., Bengio, Y. & Courville A. Deep Learning. MIT Press. (2016).

12. Bernal, J. Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge. IEEE Transactions on Medical Imaging. 36, 1231-1249 (2017).

13. Singh, H. et al. The Reduction in Colorectal Cancer Mortality After Colonoscopy Varies by Site of the Cancer. Gastroenterology Journal. 139, 1128-1137 (2010).

14. Zhao S. et al. Magnitude, Risk Factors, and Factors Associated With Adenoma Miss Rate of Tandem Colonoscopy: A Systematic Review and Meta-analysis. Gastroenterology Journal. 156, 1661-1674 (2019).

15. Castaneda, D., Popov, V.B., Verheyen, E., Wander, P. & Gross, S.A. New technologies improve adenoma detection rate, adenoma miss rate, and polyp detection rate: a systematic review and meta-analysis. Gastrointestinal Endoscopy. 88, 209-222 (2018).

16. Sánchez-Peralta, L.F., Bote-Curiel, L., Picón, A., Sánchez-Margallo, F.M. & Pagador, J.B. Deep learning to find colorectal polyps in colonoscopy: A systematic literature review. Artificial Intelligence in Medicine. 108 (2020).

17. Wang, P. et al. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nature Biomedical Engineering. 2, 741–748 (2018).

18. Bernal, J., Sánchez, J. & Vilariño, F. Towards automatic polyp detection with a polyp appearance model. Pattern Recognition. 45, 3166-3182 (2012).

19. Vazquez, D. et al. A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images. Journal of Healthcare Engineering. 2017 (2017).

20. Yuan, Z. et al. Automatic polyp detection in colonoscopy videos. SPIE Medical Imaging. 1, 1-10 (2017).

21. Mo, X., Tao, K., Wang, Q. & Wang, G. An efficient approach for polyps detection in endoscopic videos based on faster R-CNN. International Conference on Pattern Recognition. 1, 3929-3934 (2018).

22. Lee, J.Y et al. Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets. Scientific Reports. 10 (2020).

23. Spadaccini, M. et al. Computer-aided detection versus advanced imaging for detection of colorectal neoplasia: a systematic review and network meta-analysis. Lancet Gastroenterology & Hepatology. 6, 794-802 (2021).

24. Hussein, M. et al. A new artificial intelligence system successfully detects and localises early neoplasia in Barrett's esophagus by using convolutional neural networks. United European Gastroenterol Journal. (2022).

25. Hou, W. et al. Early neoplasia identification in Barrett’s esophagus via attentive hierarchical aggregation and self-distillation. Medical Image Analysis. 72 (2021).

26. Wallace, M.B. et al. Impact of Artificial Intelligence on Miss Rate of Colorectal Neoplasia. Gastroenterology. (2022).

27. Van Berkel, N. et al. Initial Responses to False Positives in AI-Supported Continuous Interactions: A Colonoscopy Case Study. ACM Transactions on Interactive Intelligent Systems. 12, 1-18 (2022).

28. Pannala, R. et al. Artificial intelligence in gastrointestinal endoscopy. VideoGIE. 5, 598-613 (2020).

29. Singh, D. & Singh, B. Effective and efficient classification of gastrointestinal lesions: combining data preprocessing, feature weighting, and improved ant lion optimization. Journal of Ambient Intelligence and Humanized Computing. 12, 8683–8698 (2021).

30. Mesejo, P. et al. Computer-aided classification of gastrointestinal lesions in regular colonoscopy. IEEE Transactions on Medical Imaging. 35, 2051–2063 (2016).

31. Byrne, M.F. et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut. 68, 94-100 (2019).

32. Patel, K. et al. A comparative study on polyp classification using convolutional neural networks. Plos One. 15, 1-16 (2020).

33. Li, K. et al. Colonoscopy Polyp Detection and Classification: Dataset Creation and Comparative Evaluations. Plos One. 16, 1-26 (2021).

34. Nogueira-Rodríguez, A. et al. Deep Neural Networks approaches for detecting and classifying colorectal polyps. Neurocomputing. 423, 721-734 (2021).

35. Ozawa, T. Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. Therapeutic Advances in Gastroenterology. 13 (2020).

36. Pu, L.Z.C.T. et al. Randomised controlled trial comparing modified Sano's and narrow band imaging international colorectal endoscopic classifications for colorectal lesions. World journal of gastrointestinal endoscopy. 10, 210-218 (2018).

37. Zorron Cheng Tao Pu, L. et al. Computer-aided diagnosis for characterization of colorectal lesions: comprehensive software that includes differentiation of serrated lesions. Gastrointestinal Endoscopy. 92, 891-899 (2020).

38. Takeda, K. et al. Accuracy of diagnosing invasive colorectal cancer using computer-aided endocytoscopy. Endoscopy. 49, 798-802 (2017).

39. Ito, N. et al. Endoscopic diagnostic support system for cT1b colorectal cancer using deep learning. Oncology. 96, 44-50 (2019).

40. Endoscopic Classification Review Group. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. Endoscopy. 37, 570-578 (2005).

41. Kominami, Y. et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. Gastrointestinal Endoscopy. 83, 643-649 (2016).

42. Jin, E.H. et al. Improved Accuracy in Optical Diagnosis of Colorectal Polyps Using Convolutional Neural Networks with Visual Explanations. Gastroenterology. 158, 2169-2179 (2020).

43. Freedman, D. et al. Detecting Deficient Coverage in Colonoscopies. IEEE Transactions on Medical Imaging. 39, 3451-3462 (2020).

44. Hartley, R. & Zisserman, A. Multiple View Geometry in Computer Vision (2nd. ed.). Cambridge University Press. (2003).

45. Mur-Artal, R., Montiel, J.M.M. & Tardos, J.D. ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE transactions on robotics. 31, 1147-1163 (2015).

46. Mahmoud, N. et al. ORBSLAM-Based Endoscope Tracking and 3D Reconstruction. Computer-Assisted and Robotic Endoscopy. International Workshop on Computer-Assisted and Robotic Endoscopy. 1, 72-83 (2017).

47. Mahmoud, N. et al. SLAM based quasi dense reconstruction for minimally invasive surgery scenes. arXiv preprint. (2017).

48. Mahmoud, N. et al. Live Tracking and Dense Reconstruction for Handheld Monocular Endoscopy. IEEE Transactions on Medical Imaging. 38, 79-89 (2019).

49. Docea, R. et al. Simultaneous localisation and mapping for laparoscopic liver navigation: a comparative evaluation study. Medical Imaging. 11598, 62-76 (2021).

50. Parashar, S., Pizarro, D. & Bartoli, A. Isometric non-rigid shape-from motion with Riemannian geometry solved in linear time. EEE Transactions on Pattern Analysis and Machine Intelligence. 40, 2442–2454 (2018).

51. Lamarca, J., Parashar, S., Bartoli, A. & Montiel, J.M.M. DefSLAM: Tracking and Mapping of Deforming Scenes from Monocular Sequences. IEEE Transactions on robotics. 37, 291-303 (2020).

52. Rodriguez, J.J.G., Lamarca, J., Morlana, J., Tardos J.D. & Montiel, J.M.M. Sd-defslam: Semi-direct monocular slam for deformable and intracorporeal scenes. arXiv preprint. (2020).

53. Sengupta, A. & Bartoli, A. Colonoscopic 3D reconstruction by tubular non-rigid structure-from-motion. International Journal of Computer Assisted Radiology and Surgery. 16, 1237–1241 (2021).

54. Lin, J. et al. Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks. Medical Image Analysis. 48, 162-176. (2018).

55. Ma, R. et al. RNNSLAM: Reconstructing the 3D colon to visualize missing regions during a colonoscopy. Medical Image Analysis. 72 (2021).

56. Recasens, D., Lamarca, J., Facil, J.M., Montiel, J.M.M. & Civera, J. Endo-Depth-and-Motion: Reconstruction and Tracking in Endoscopic Videos using Depth Networks and Photometric Constraints. IEEE Robotics and Automation Letters. 6, 7225-7232 (2021).

57. Zhang, S., Zhao, L., Huang, S., Ye, M. & Hao, Q. A Template-Based 3D Reconstruction of Colon Structures and Textures From Stereo Colonoscopic Images. IEEE Transactions on Medical Robotics and Bionics. 3, 85-95 (2021).

58. Rau, A., Bhattarai, B., Agapito, L. & Stoyanov, D. Bimodal Camera Pose Prediction for Endoscopy. arXiv preprint. (2022).

59. Takiyama, H. et al. Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks. Nature Scientific Reports. 8 (2018).

60. Igarashi, S., Sasaki, Y., Mikami, T., Sakuraba, H. & Fukuda, S. Anatomical classification of upper gastrointestinal organs under various image capture conditions using AlexNet. Computers in Biology and Medicine. 124 (2020).

61. Beg, S. et al. Quality standards in upper gastrointestinal endoscopy: a position statement of the British Society of Gastroenterology (BSG) and Association of Upper Gastrointestinal Surgeons of Great Britain and Ireland (AUGIS). Gut. 66, 1886–1899 (2017).

62. Rey, J.F. & Lambert, R. The ESGE Quality Assurance Committee: ESGE recommendations for quality control in gastrointestinal endoscopy: guidelines for image documentation in upper and lower GI endoscopy. Endoscopy. 33, 901-903 (2001).

63. Yao, K. The endoscopic diagnosis of early gastric cancer. Annals of Gastroenterology. 26, 11-22 (2013).

64. He, Q. et al. Deep Learning Based Anatomical Site Classification for Upper Gastrointestinal Endoscopy. International Journal of Computer Assisted Radiology and Surgery. 15, 1085-1094 (2020).

65. Jha, D. et al. A comprehensive analysis of classification methods in gastrointestinal endoscopy imaging. Medical Image Analysis. 70 (2021).

66. Riegler, M. et al. Multimedia for medicine: the medico task at mediaeval 2017. CEUR Workshop Proceedings - Multimedia Benchmark Workshop. (2017).

67. Pogorelov, K. et al. Medico multimedia task at mediaeval 2018. CEUR Workshop Proceedings. (2018).

68. Hicks, S.A. et al. ACM Multimedia BioMedia 2020 Grand Challenge Overview. ACM International Conference on Multimedia. 1, 4655-4658 (2020).

69. Pogorelov, K. et al. KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. ACM International Conference on Multimedia. 1, 164–169 (2017).

70. Pogorelov, K. Nerthus: A Bowel Preparation Quality Video Dataset. ACM International Conference on Multimedia. 1, 170-174 (2017).

71. Luo, Z., Wang, X., Xu, Z., Li, X. & Li, J. Adaptive Ensemble: Solution to the Biomedia ACM MM GrandChallenge 2019. ACM International Conference on Multimedia. 1, 2583-2587(2019).

72. Saito, H. et al. Automatic anatomical classification of colonoscopic images using deep convolutional neural networks. Gastroenterology Report. 9, 226–233 (2021).

73. Sestini, L., Rosa, B., De Momi, E., Ferrigno, G. & Padoy, N. A Kinematic Bottleneck Approach for Pose Regression of Flexible Surgical Instruments Directly From Images. IEEE Robotics and Automation Letters. 6, 2938-2945 (2021).

74. Roß, T. et al. Comparative validation of multi-instance instrument segmentation in endoscopy: Results of the ROBUST-MIS 2019 challenge. Medical Image Analysis. 70 (2021).

75. Gonzalez, C., Bravo-Sanchez, L. & Arbelaez, P. ISINet: An Instance-Based Approach for Surgical Instrument Segmentation. Medical Image Computing and Computer Assisted Intervention. 12263, 595-605 (2020).

76. Xiaowen, K. et al. Accurate instance segmentation of surgical instruments in robotic surgery: model refinement and cross-dataset evaluation. International Journal of Computer Assisted Radiology and Surgery. 19, 1607-1614 (2021).

77. Colleoni, E., Edwards, P. & Stoyanov, D. Synthetic and Real Inputs for Tool Segmentation in Robotic Surgery. Medical Image Computing and Computer Assisted Intervention. 12263, 700-710 (2020).

78. Colleoni, E. & Stoyanov, D. Robotic Instrument Segmentation With Image-to-Image Translation. IEEE Robotics and Automation Letters. 6, 935-942 (2021).

79. Pfeiffer, M. et al. Generating large labeled data sets for laparoscopic image processing tasks using unpaired image-to-image translation. Medical Image Computing and Computer Assisted Intervention. 11768, 119-127 (2019).

80. Sahu, M., Stromsdorfer, R., Mukhopadhyay, A. & Zachow, S. Endo-Sim2Real: Consistency Learning-Based Domain Adaptation for Instrument Segmentation. Medical Image Computing and Computer Assisted Intervention. 12263, 784-794 (2020).

81. Zhang, Z., Rosa, B. & Nageotte, F. Surgical Tool Segmentation Using Generative Adversarial Networks With Unpaired Training Data. IEEE Robotics and Automation Letters. 6, 6266-6273 (2021).

82. Zhao, Z. et al. One to Many: Adaptive Instrument Segmentation via Meta Learning and Dynamic Online Adaptation in Robotic Surgical Video. IEEE International Conference on Robotics and Automation. 1, 13553-13559 (2021).

83. Du, X. et al. Articulated Multi-Instrument 2-D Pose Estimation Using Fully Convolutional Networks. IEEE Transactions on Medical Imaging. 37, 1276-1287 (2018).

84. Kayhan, M. et al. Deep Attention Based Semi-supervised 2D-Pose Estimation for Surgical Instruments. Pattern Recognition, ICPR International Workshops and Challenges. 1, 444-460 (2021).

85. Allan, M., Ourselin, S., Hawkes, D.J., Kelly, J.D. & Stoyanov, D. 3-D Pose Estimation of Articulated Instruments in Robotic Minimally Invasive Surgery. IEEE Transactions on Medical Imaging. 37, 1204-1213 (2018).

86. Hasan, K., Calvet, L., Rabbani, N. & Bartoli, A. Detection, segmentation, and 3D pose estimation of surgical tools using convolutional neural networks and algebraic geometry. Medical Image Analysis. 70 (2021).

87. Ahmidi, N. et al. A Dataset and Benchmarks for Segmentation and Recognition of Gestures in Robotic Surgery. IEEE Transactions on Biomedical Engineering. 64, 2025-2041 (2017).

88. Van Amsterdam, B., Clarkson, M.J. & Stoyanov, D. Gesture Recognition in Robotic Surgery: A Review. IEEE Transactions on Biomedical Engineering. 68, 2021-2035 (2021).

89. Dergachyova, O., Bouget, D., Huaulmé, A., Morandi, X. & Jannin, P. Automatic data-driven real-time segmentation and recognition of surgical workflow. International Journal of Computer Assisted Radiology and Surgery. 11, 1081-1090 (2016).

90. Lalys, F. & Jannin, P. Surgical process modelling: A review. International Journal of Computer Assisted Radiology and Surgery. 9, 495–511 (2014).

91. Oleari, E. et al. Enhancing Surgical Process Modeling for Artificial Intelligence development in robotics: the SARAS case study for Minimally Invasive Procedures. International Symposium on Medical Information and Communication Technology. 1, 1-6 (2019).

92. Gurcan, I. & Van Nguyen, H. Surgical activities recognition using multi-scale recurrent networks. IEEE International Conference on Acoustics, Speech and Signal Processing. 1, 2887-2891 (2019).

93. Funke, I. et al. Video-based surgical skill assessment using 3D convolutional neural networks. *International Journal of Computer Assisted Radiology and Surgery. 14***,** 1217–1225 (2019).

94. Qin, Y., Allan, M., Burdick, J.W. & Azizian, M. Autonomous Hierarchical Surgical State Estimation During Robot-Assisted Surgery Through Deep Neural Networks. IEEE Robotics and Automation Letters. 6, 6220-6227 (2021).

95. Park, J. & Park, C.H. Recognition and Prediction of Surgical Actions Based on Online Robotic Tool Detection. IEEE Robotics and Automation Letters. 6, 2365-2372 (2021).

96. Long, Y. et al. Relational Graph Learning on Visual and Kinematics Embeddings for Accurate Gesture Recognition in Robotic Surgery. IEEE International Conference on Robotics and Automation. 1, 13346-13353 (2021).

97. Van Amsterdam, B. et al. Gesture Recognition in Robotic Surgery with Multimodal Attention. IEEE Transactions on Medical Imaging. (2022).

98. Stauder, R. et al. The TUM LapChole dataset for the M2CAI 2016 workflow challenge. arXiv preprint. (2016).

99. Twinanda, A.P. et al. MICCAI Modeling and Monitoring of Computer Assisted Interventions Challenge. (2016).

100. Twinanda, A.P. et al. EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos. IEEE Transactions on Medical Imaging. 36, 86-97 (2017).

101. Jin, Y. et al. Multi-task recurrent convolutional network with correlation loss for surgical video analysis. Medical Image Analysis. 59 (2020).

102. Bawa, V.S. et al. ESAD: Endoscopic Surgeon Action Detection Dataset. arXiv preprint. (2021).

103. Bawa, V.S. et al. The SARAS Endoscopic Surgeon Action Detection (ESAD) dataset: Challenges and methods. arXiv preprint. (2021).

104. Kitaguchi, D. et al. Automated laparoscopic colorectal surgery workflow recognition using artificial intelligence: Experimental research. International Journal of Surgery. 79, 88-94 (2020).

105. Ban, Y. et al. SUrgical PRediction GAN for Events Anticipation. arXiv preprint. (2021).

106. Nwoye, C.I. et al. Rendezvous: Attention Mechanisms for the Recognition of Surgical Action Triplets in Endoscopic Videos. Medical Image Analysis. 78 (2022).

107. Nwoye, C.I. et al. CholecTriplet2021: a benchmark challenge for surgical action triplet recognition. arXiv PrePrint. (2022).

108. Gibaud, B. Toward a standard ontology of surgical process models. International Journal of Computer Assisted Radiology and Surgery. 13, 1397-1408 (2018).

109. Katić, D. et al. LapOntoSPM: an ontology for laparoscopic surgeries and its application to surgical phase recognition. International Journal of Computer Assisted Radiology and Surgery. 10, 1427-1434 (2015).

110. Mascagni, P. & Padoy, N. OR black box and surgical control tower: Recording and streaming data and analytics to improve surgical care. Journal of Visceral Surgery. 158, 18-25 (2021).

111. Funke, I., Mees, S.T., Weitz, J. & Speidel, S. Video-based surgical skill assessment using 3D convolutional neural networks. International Journal of Computer Assisted Radiology and Surgery. 14, 1217-1225 (2019).

112. Wang, T., Wang, Y. & Li, M. Towards Accurate and Interpretable Surgical Skill Assessment: A Video-Based Method Incorporating Recognized Surgical Gestures and Skill Levels. Medical Image Computing and Computer Assisted Intervention. 12263 (2020).

113. Collins, J.W. et al. Ethical implications of AI in robotic surgical training: A Delphi consensus statement. European Urology Focus. 8, 613-622 (2022).

114. Lavanchy, J.L. et al. Automation of surgical skill assessment using a three-stage machine learning algorithm. Nature Science Reports. 11 (2021).

115. Liu, D. et al. Towards Unified Surgical Skill Assessment. IEEE/CVF Conference on Computer Vision and Pattern Recognition. 1, 9522-9531 (2021).

116. Vedula, S.S. et al. Artificial Intelligence Methods and Artificial Intelligence-Enabled Metrics for Surgical Education: A Multidisciplinary Consensus. Journal of the American College of Surgeons. 234, 1181-1192 (2022).

117. Zhu, Y., Xu, Y., Chen, W., Zhao, T. & Zheng S. A CNN-based Cleanliness Evaluation for Bowel Preparation in Colonoscopy. International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. 1, 1-5 (2019).

118. Hutchinson, K., Li, Z., Cantrell, L.A., Schenkman, N.S. & Alemzadeh, H. Analysis of Executional and Procedural Errors in Dry-lab Robotic Surgery Experiments. arXiv preprint. (2021).

119. Zia, A. et al. Endoscopic Vision Challenge 2022, 25th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2022). (2022).

120. Shademan, A. et al. Supervised autonomous robotic soft tissue surgery. Science translational medicine. 8 (2016).

121. Saeidi, H. et al. Autonomous robotic laparoscopic surgery for intestinal anastomosis. Science Robotics. 7 (2022).

122. Dehghani, H. & Kim, P.C.W. Robotic Automation for Surgery. Digital Surgery. 1, 203-213 (2021).

123. Oberlin, J., Buharin, V.E., Dehghani, H. & Kim, P.C.W. Intelligence and Autonomy in Future Robotic Surgery. Robotic Surgery. 1, 183-195 (2021).

124. Kassahun, Y. et al. Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. International Journal of Computer Assisted Radiology and Surgery. 11, 553–568 (2016).

125. Haidegger, T. Autonomy for Surgical Robots: Concepts and Paradigms. IEEE Transactions on Medical Robotics and Bionics. 1, 65-76 (2019).

126. Attanasio, A.A., Scaglioni, B., De Momi, E., Fiorini, P. & Valdastri, P. Autonomy in Surgical Robotics. Annual Review of Control, Robotics, and Autonomous Systems. 4, 651-679 (2021).

127. 108. Varier, V.M. et al. Collaborative Suturing: A Reinforcement Learning Approach to Automate Hand-off Task in Suturing for Surgical Robots. IEEE International Conference on Robot and Human Interactive Communication. 1, 1380-1386 (2020).

128. Nguyen, T., Nguyen, N.D., Bello, F. & Nahavandi, S. A New Tensioning Method using Deep Reinforcement Learning for Surgical Pattern Cutting. IEEE International Conference on Industrial Technology. 1, 1339-1344 (2019).

129. Attanasio, A. et al. Autonomous Tissue Retraction in Robotic Assisted Minimally Invasive Surgery – A Feasibility Study. IEEE Robotics and Automation Letters. 5, 6528-6535 (2020).

130. Gruijthuijsen, C. et al. Autonomous Robotic Endoscope Control based on Semantically Rich Instructions. arXiv preprint. (2021).

131. Shin, C. Autonomous Tissue Manipulation via Surgical Robot Using Learning Based Model Predictive Control. International Conference on Robotics and Automation. 1, 3875-3881 (2019).

132. Omisore, O.M. et al. A Review on Flexible Robotic Systems for Minimally Invasive Surgery. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 52, 631-644 (2022).

133. Martin, J.W. et al. Enabling the future of colonoscopy with intelligent and autonomous magnetic manipulation. Nature Machine Intelligence. 2, 595-606 (2020).

134. Loftus, T.J. et al. Intelligent, Autonomous Machines in Surgery. Journal of Surgical Research. 253, 92-99 (2020).

135. Hung, A.J., Chen, J. & Gill, I.S. Automated Performance Metrics and Machine Learning Algorithms to Measure Surgeon Performance and Anticipate Clinical Outcomes in Robotic Surgery. JAMA Surgery. 153, 770–771 (2018).

136. Ahmad, O.F., Stoyanov, D. & Lovat, L.B. Barriers and pitfalls for artificial intelligence in gastroenterology: Ethical and regulatory issues. Techniques and Innovations in Gastrointestinal Endoscopy. 22, 80-84 (2020).

137. Muehlematter, U.J., Daniore, P. & Vokinger, K.N. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015-20): a comparative analysis. Lancet Digit Health. 3, 195-203 (2021).

138. Taghiakbari, M., Mori, Y. & von Renteln, D. Artificial intelligence-assisted colonoscopy: A review of current state of practice and research. World Journal of Gastroenterology. 27, 8103-8122 (2021).

139. Vulpoi, R-A. et al. Artificial Intelligence in Digestive Endoscopy—Where Are We and Where Are We Going? Diagnostics. 12 2022.

140. Mori, Y., Bretthauer, M. & Kalager, M. Hopes and hypes for artificial intelligence in colorectal cancer screening. Gastroenterology. 161, 774-777 (2021).

141. Aisu, N. et al. Regulatory-approved deep learning/machine learning-based medical devices in Japan as of 2020: A systematic review. PLOS Digital Health. (2022).

142. Mori, Y., Neumann, H., Misawa, M., Kudo, S. & Bretthauer, M. Artificial intelligence in colonoscopy-Now on the market. What's next?. Journal of Gastroenterology and Hepatology. 36, 7-11 (2021).

143. Areia, M. et al. Cost-effectiveness of artificial intelligence for screening colonoscopy: a modelling study. The Lancet Digital Health. (2022).

144. The MONAI Consortium. Project MONAI. Zenodo. (2020).

Acknowledgements:

This work was supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) at the University College London (203145Z/16/Z), EPSRC (EP/P012841/1, EP/P027938/1, and EP/R004080/1), and the H2020 FET (GA 863146). D. Stoyanov is supported by a Royal Academy of Engineering Chair in Emerging Technologies (CiET1819\2\36) and an EPSRC Early Career Research Fellowship (EP/P012841/1).

Competing Interests:

D. Stoyanov is part of Digital Surgery from Medtronic plc. and a shareholder in Odin Vision Ltd. L.B. Lovat is a shareholder in Odin Vision Ltd. The authors have no conflict of interests to declare.

Proposed display items:

Boxes: 0

Figures: 5

Tables: 0

Other: 0

Keywords: Artificial intelligence; Machine learning; Endoscopy; Ethics; Regulation

Key references: Artificial Intelligence, Computer-Assisted Diagnosis, Computer-Assisted Intervention, Robotically Assisted Surgery.

Figure legends:

Figure 1: AI-based CAI in gastrointestinal MIS. AI has become a key component to the development of computer-assisted intervention in gastrointestinal endoscopy.

Figure 2: AI-based CADe and CADx systems. CNNs can efficiently detect and diagnose gastrointestinal lesions perioperatively. CADe systems significantly alleviate the miss-rate of adenomas in gastrointestinal endoscopy.

Figure 3: AI-based computer-assisted navigation. CNN-based SLAM pipelines avoid the need for tracking visual features by directly estimating the depth map of the endoscopic scene from a single view. Despite noteworthy progress, Computer-assisted surgical navigation remains challenging in gastrointestinal endoscopy.

Figure 4: AI-based surgical workflow recognition. Surgical workflow recognition is a key to the automation of endoscopy reporting and surgical procedures. Temporal NNs demonstrated promising results for recognizing surgical activities in gastrointestinal endoscopy.

Figure 5: AI-based robotically-assisted surgery. The combined advances of surgical robotics and AI opened up promising perspectives for the development of robotically-assisted surgery. Nevertheless, efficiently combining heterogeneous sensor data remains a major impediment to the automation of MIS procedures in gastroenterology.