UNIVERSITÁ DEGLI STUDI DI NAPOLI
“FEDERICO II”
Scuola di Medicina e Chirurgia

DOTTORATO DI RICERCA XXXV CICLO
DIPARTIMENTO DI SANITA’ PUBBLICA E MEDICINA PREVENTIVA
Direttrice: Chiar.$^{ma}$ Prof. Maria Triassi

TESI DI DOTTORATO

# “COLORECTAL 100”

*Application of Artificial Intelligence and Computer Vision in Laparoscopic Colorectal Surgery*

**Dottoranda**
Dott.ssa Ludovica Guerriero

**Relatore**
Chiar.mo Prof. Francesco Corcione

**Correlatrice**
Chiar.ma Maria Triassi

*Anno Accademico 2023-2024*

**INDEX**

## 1. INTRODUCTION

The effectiveness and volume of surgery has dramatically increased in the last decades, with more than 300 million procedures performed in 2012 [1]. However, surgery is still not the safe place we would like it to be, as surgical adverse events represent a great part of medical mistakes [2]. On this basis, surgical safety can be now considered a research priority.

When performing an operation, surgeons need to communicate with a team of collaborators, interpret multiple signals coming from screens and other devices, project surgical principles into the present case, anticipate consequences of decisions and act in a timely manner. The success and coordination of all these complex events results in good quality surgery. This process demands a high physical and cognitive effort and any minimum mistake can lead to consequences for patients.

In this thesis key elements enabling the vision of data-driven solutions in surgery will be shown and a future scenario in which advanced analytics are used to promote safety in laparoscopic colorectal surgery will be proposed with a perspective on the future goals we expect to achieve.

The *Colorectal 100* project was born in collaboration with the I-Cube group from Strasbourg (University of Strasbourg, IHU Strasbourg, France) with the idea of creating a dataset of 50 videos of Laparoscopic Left Hemicolectomy (LLH) and 50 videos of Laparoscopic Right Hemicolectomy (LRH) to be analyzed with a specific AI tool box to obtain automatic phases and steps recognition. This "smart library" for colorectal surgery will serve as a basis for future research such as automatic anatomical recognition and multicentric studies for external validation with other European tertiary center surgical departments. Artificial intelligence (AI) can be loosely defined as the study of algorithms that give machines the ability to reason and perform cognitive functions such as problem solving, object and word recognition, and decision-making. AI has increasingly become the topic of both popular and academic literature as years of research have finally built to thresholds of knowledge that have rapidly generated practical applications. However, as with many emerging technologies, the true promise of AI can be lost in its hype and lead to useless results for clinical practice. It is, therefore, important for surgeons to have a foundation of knowledge of AI to understand how it may impact healthcare and to consider ways in which they may interact with this technology.

AI's roots are found across multiple fields, including robotics, philosophy, psychology, linguistics, and statistics. Major advances in computer science, such as improvements in processing speed and power, have functioned as a catalyst to allow for the base technologies required for the advent of AI. The growing popularity of AI across many different industries has attracted venture capital investment up to $5 billion in 2016 alone [3]. Much of the current attention on AI has focused on the four core subfields introduced below.

### 1.1 Machine Learning

Machine learning (ML) enables machines to learn and make predictions by recognizing patterns. Traditional computer programs are explicitly programmed with a desired behavior. ML allows a computer to utilize partial labeling of the data (supervised learning) or the structure detected in the data itself (unsupervised learning) to explain or make predictions about the data without explicit programming. Supervised learning is useful for training a ML algorithm to predict a known result or outcome while unsupervised learning is useful in searching for patterns within data. A third category within machine learning is reinforcement learning, where a program attempts to accomplish a task (e.g. inferring medical decisions) while learning from its own successes and mistakes. ML is particularly useful for identifying subtle patterns in large datasets – patterns that may be imperceptible to humans performing manual analyses – by employing techniques that allow for more indirect and complex non-linear relationships and multivariate effects than conventional statistical analysis. ML has outperformed logistic regression for prediction of surgical site infections (SSI) by building non-linear models that incorporate multiple data sources, including diagnoses, treatments, and laboratory values. Furthermore, multiple algorithms working together (ensemble ML) can be used to calculate predictions at accuracy levels thought to be unattainable with conventional statistics. For example, by analyzing patterns of diagnostic and therapeutic data (including surgical resection) in the Surveillance, Epidemiology and End Results (SEER) cancer registry and comparing data to Medicare claims, ensemble ML with random forests, neural networks, and lasso regression was able to predict patient lung cancer staging by using International Classification of Diseases (ICD)-9 claims data alone with 93% sensitivity, 92% specificity, and 93% accuracy, outperforming a decision tree approach based on clinical guidelines alone (53% sensitivity, 89% specificity, 72% accuracy)[4].

### 1.2 Natural Language Processing

Natural language processing (NLP) is a subfield that emphasizes building a computer's ability to understand human language and is crucial for large scale analyses of content such as electronic medical record (EMR) data, especially physicians' narrative documentation. To achieve human-level understanding of language, successful NLP systems must expand beyond simple word recognition to incorporate semantics and syntax into their analyses. NLP allows clinicians to write more naturally rather than having to input specific text sequences or select from menus to allow a computer to recognize the data. In surgical patients, NLP has been used to automatically comb through EMRs to identify words and phrases in operative reports and progress notes that predicted anastomotic leak after colorectal resections. Many of its predictions reflected simple clinical knowledge that a surgeon would have (e.g. operation type and difficulty), but the algorithm was also able to adjust predictive weights of phrases describing patients (e.g. irritated, tired) relative to the postoperative day to achieve predictions of leak with a sensitivity of 100% and specificity of 72%. [5]. The ability of algorithms to self-correct can increase the utility of their predictions as datasets grow to become more representative of a patient population.

### 1.3 Computer Vision

Computer Vision (CV) describes machine understanding of images and videos, and significant advances have resulted in machines achieving human-level capabilities in areas such as object and scene recognition. Important healthcare-related work in computer vision includes image acquisition and interpretation in axial imaging with applications including computer-aided diagnosis, image-guided surgery, and virtual colonoscopy. Initially influenced by statistical signal processing, the field has recently shifted significantly towards more data-intensive ML approaches, such as neural networks, with adaptation into new applications. For example, real-time analysis of laparoscopic video has yielded 92.8% accuracy in automated identification of the steps of a sleeve gastrectomy and noted missing or unexpected steps. While predictive video analysis is in its infancy, such work provides proof-of-concept that AI can be leveraged to process massive amounts of surgical data to identify or predict adverse events in real-time for intraoperative clinical decision support (Figure 1).

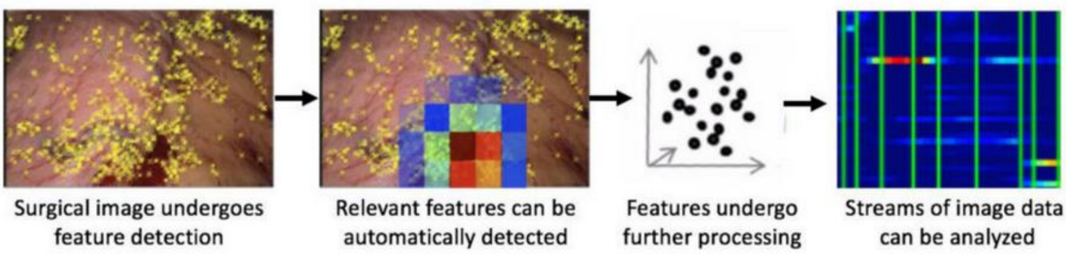


**Figure 1**. Computer vision utilizes mathematical techniques to analyze visual images or video streams as quantifiable features such as color, texture, and position that can then be used within dataset to identify statistically meaningful events such as bleeding.

### 1.4 Artificial Neural Networks

Artificial neural networks, a subfield of ML, are inspired by biological nervous systems and have become of paramount importance in many AI applications. Neural networks process signals in layers of simple computational units (neurons); connections between neurons are then parameterized via weights that change as the network learns different input-output maps corresponding to tasks such as pattern/image recognition and data classification (Figure 1**)**. Deep learning networks are neural networks composed of many layers and are able to learn more complex and subtle patterns than simple one or two-layer neural networks. Clinically, ANNs have significantly outperformed more traditional risk prediction approaches. For example, an ANN's sensitivity (89%) and specificity (96%) outperformed APACHE II sensitivity (80%) and specificity (85%) for prediction of pancreatitis severity six hours after admission**.** [6]. By using clinical variables such as patient history, medications, blood pressure, and length of stay, ANNs, in combination with other ML approaches, have yielded predictions of in-hospital mortality after open abdominal aortic aneurysm repair with sensitivity of 87%, specificity of 96.1%, and accuracy of 95.4% [7].

### 1.5 Clinical applications

Clinical applications of such work include being used to support surgical practice. The successful utilization of deep learning to create a computer vision algorithm for the classification of benign and malignant skin lesions at an accuracy level equivalent to dermatologists [8].

NLP and ML analyses of postoperative colorectal patients demonstrated that prediction of anastomotic leaks improved to 92% accuracy when different data types were analyzed in concert instead of individually (accuracy of vital signs – 65%; lab values – 74%; text data – 83%) [9].

For AI, much of its clinical potential is in its ability to analyze combinations of structured and unstructured data (e.g. EMR notes, vitals, laboratory values, video, and other aspects of "big data") to generate clinical decision support. Each type of data could be analyzed independently or in concert with different types of algorithms to yield innovations.
The true potential of AI remains to be seen and could be difficult to predict at this time. Synergistic reactions between different technologies can lead to revolutionary technology; for example, recent synergistic combinations of advanced robotics, computer vision, and neural networks led to the advent of autonomous cars. Similarly, independent components within AI and other fields could combine to create changes to healthcare delivery. Surgeons should be engaged in assessing the applicability of AI advances to ensure appropriate translation to clinical practice.

**1.6 Big data and artificial intelligence for Surgical Data Science**

The growing uptake of image-guided interventions, such as minimally invasive surgery, interventional radiology and surgical endoscopy, is changing the way we interpret surgery. Indeed, the images guiding these procedures, whether radiological or endoscopic, are a natural source of direct, unbiased and rich information on intraoperative events. These digital images are much more informative and reliable than operator dictated post-operative reports. The analysis of videos of surgical procedures and OR activities could offer strategies to improve surgical care. This is especially true for procedures performed with a minimally invasive approach. In fact, in laparoscopic surgery the partial loss of haptic feedback is compensated by magnified, high-definition videos acquired by endoscopic cameras. Endoscopic videos guiding surgical procedures represent a direct and readily available source of digital data on the intraoperative phase of surgical care. In recent years, the analysis of endoscopic videos of minimally invasive surgical procedures has enabled the study of the impact of OR activities on patient outcomes [10] and the assessment of quality improvement initiatives. In addition, video-based assessment is being increasingly investigated for operative performance assessment, formative feedback, and surgical credentialing. However, it has mostly remained confined to the research domain given the burden of manually reviewing and consistently assessing surgical videos [11]. In the *RightHemicol50* dataset we examined videos of laparoscopic right hemicolectomy and we discussed the application of automated video analysis in laparoscopic colorectal surgery tracing the way to possible paths towards the clinical value of computer vision in this specific field of surgery.  We also discuss the challenges and obstacles that remain to be overcome for broader implementation and adoption of CV in surgery.

## 2. MATERIALS AND METHODS

While evidence on the clinical value of AI-based solutions for the screening and staging of colorectal cancer (CRC) is mounting, Computer Vision and Artificial Intelligence applications to enhance the surgical treatment of CRC are still in their early stage. This study introduces key AI concepts to a surgical audience and illustrates fundamental steps to develop CV for surgical applications. 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), suggesting computer vision as a potentially valuable tool for intraoperative decision-making and guidance.

In the *Colorectal100* study the annotation process of the 100 colorectal laparoscopic videos is ongoing using the 2.8.2 version of MOSaiC annotation platform, a cloud-based collaborative video annotation platform from I-Cube, Strasbourg, according to a specific annotation protocol. The datasets are divided in *RightHemicol50* dataset, including 50 videos of standardized laparoscopic right hemicolectomy (LRH), and *LeftHemicol50*, including 50 videos of standardized laparoscopic left hemicolectomy (LLH). All procedures were performed by an expert surgeon following a standardized technique used for all cases in the surgical school of Monaldi Hospital, Naples, Italy.

An example of annotation and recognition image is shown in Figure 2.

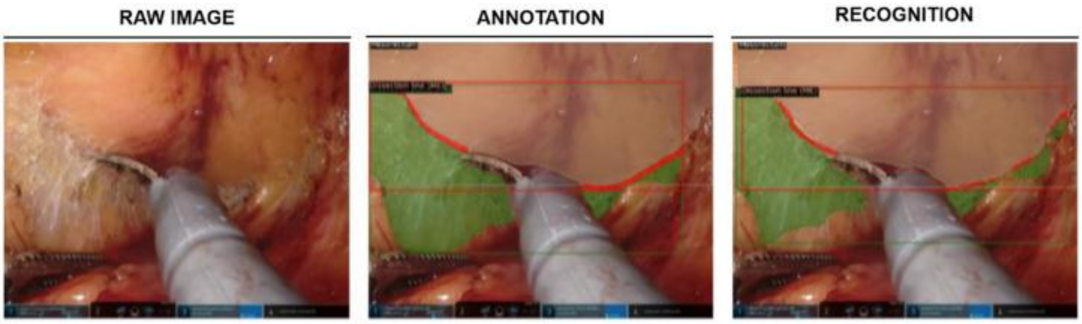


**Figure 2.** Example of annotation and recognition phase.

An entire surgical video can be classified into phases, broad stages of surgical procedures, which can be further broken down into more specific steps that are performed to achieve meaningful surgical goals such as exposing specific anatomic structures or performing anastomosis. In this series, adverse events according EAES classification [12-13] have been annotated too.

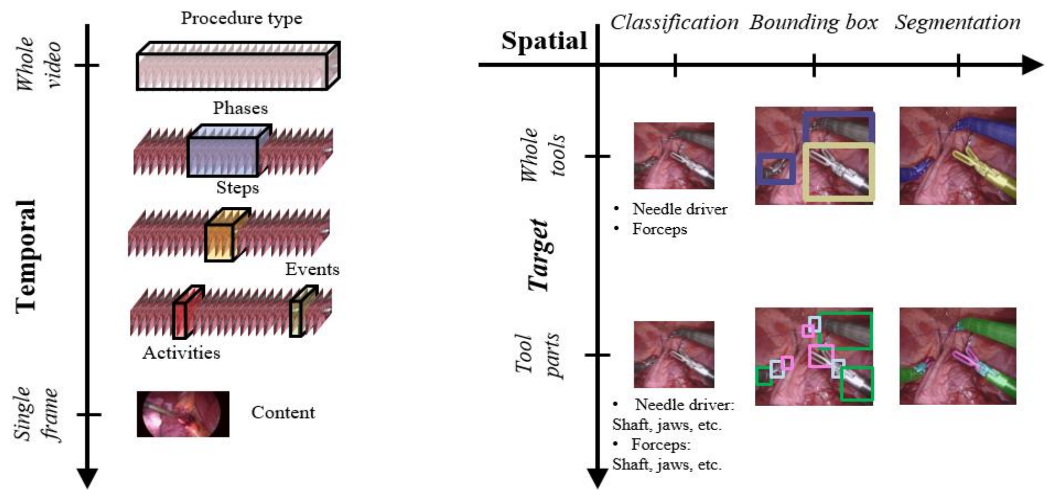


**Figure 3.** Visual representation of the division of a surgical procedure into phases and steps

## 2.1 Data collection and annotation

Fifty videos of elective of LRH procedures, for oncological purposes, performed from January 2017 to 0ctober 2023, in one Italian tertiary care center, were retrospectively analyzed. This contributing center was Monaldi Hospital, Naples. Of these, 40 videos have been uploaded on Mosaic platform and annotated in phases, steps and adverse events by a single surgeon from Monaldi Hospital.

Only complete videos of standard oncologic Right Hemicolectomy procedures performed in patients > 18 years old were included in the study. Videos not showing the beginning of the ileocolic vessel's identification and preparation or showing bailout procedures (mesocolon section first, abnormal anatomy or an intraoperative colonoscopy) were excluded from the analysis. Included LRH videos were manually annotated with an assessment of the phases and steps as reported in the annotation protocol.

Since clinical data were not collected, the local medical research and ethical committee cleared the present study from the Research Involving Human Subjects Act.

**Annotation protocol**

*Colorectal100* phase definition was designed to maximize consistency on selected procedures. The following phases and steps definition was designed to better reflect surgical semantics and to be generalizable to more complex procedures.

**RIGHT HEMICOLECTOMY "*RightHemicol50 dataset*"**

| Phase | Cues | Comments |
|---|---|---|
| Abdominal cavity exploration and preparation | - Starts when the surgeon is done placing all the trocars and the camera is steady and the surgeon starts exposing the last ileal loop | |
| Ileocolic vessels identification and preparation | - Starts when the dissecting tool is inserted, the assistant instrument (grasper) grasps the last ileal loop neck and the operator right instrument (Thunderbeat) incise the mesocolon and starts preparation of both the ileocolic vein and artery | - the dissector may be inserted for adhesiolysis before approaching the ileocolic vessels |
| Ileocolic vessels clipping and cutting | - Starts when the clipper is inserted and ends when the ileocolic vessels are dissected | |
| Steps | (s) ileocolic artery clipping<br>(s) ileocolic vein clipping | |
| Toldt-Gerota window | - Starts when the dissecting tool is re-inserted after the artery and vein are clipped and cut and the Toldt's fascia (up) is detached from the Gerota's fascia (down)<br>- Ends when the duodenum is visualized on the right of the screen and the transverse colon on the top. The right urether must be visualized as well (down) | |
| Mesocolon division | - Starts with reinsertion of the dissector and mesocolon division.<br>The right colic vessels, if present, are exposed, clipped and sectioned<br><br>- Ends when the transverse colon is | - Right colic vessels might not be present and therefore no clipping and cutting will happen |

| | | |
|---|---|---|
| | reached and its wall is visible | |
| Last jejunal loop section | - Starts with mesentery section<br>- Ends with insertion of the stapler and jejunal loop stapling | We use 1 white cartridge |
| Mesocolon division from above | - Starts with the dissection of hepatic flexure from above (we can see the liver and gallbladder on our right now)<br>-Ends with the complete mobilization of the right parietocolic gutter | |
| Transverse colon stapling | - Starts with the insertion of the stapler from the left side of the screen<br>- Ends when the transverse colon is sectioned | We use 1 blue cartridge. Sometimes 2 cartridges are needed.<br><br>A bleeding from the suture line might occur and coagulation might be necessary |
| Ileo-transverse colon anastomosis | - Starts with last jejunal loop approximation to the stapled transverse colon.<br>The stapler is inserted and fired.<br>A single stitch is placed at the inferior corner of the defect first.<br>Then the defect is closed with a double running suture (V-lock+PDS) from up to down.<br>- Ends with the knot of the PDS suture | After firing, the stapler is retracted out of the field. |
| Steps | (s) firing stapler<br>(s) angle point<br>(s) first layer<br>(s) second layer | The defect is closed in a double layer fashion: 1) V-lock running suture from up to down<br>2) PDS oversew suture |

| | | |
|---|---|---|
| Closure of the mesenteric defect | - Starts with mesenteric defect exposure and closing with a running Vicryl suture<br>- Ends with the knot of the suture | The length of the mesenteric defect may be variable |
| Right hemicolectomy packaging | - Starts when the endobag is inserted form the left of the screen<br>- Ends when the bag is closed with the colon inside | |
| Cleaning and coagulation | Possible starting points:<br>After the colon is placed in the endobag and placed on top of the liver, the camera takes a look around for final checking. The first frame that shows the abdominal cavity is the beginning of this phase.<br>Any of the following clues triggers the start of this phase, anytime:<br>- presence of the suction tool<br>- presence of the drainage<br>- presence of the gauze | |

**Table 1.** Laparoscopic Right Hemicolectomy annotation protocol, including phases and steps.

### 2.2 Laparoscopic Right Hemicolectomy Workflow Analysis

**Dataset**

Within the *Colorectal100* study, *RightHemicol50* is a dataset composed of 50 surgical videos of right hemicolectomy performed in the Laparoscopic and Robotic General Surgery Department of Monaldi Hospital, Naples, Italy. Fourty videos were uploaded on Mosaic platform. All the surgeries were performed by an expert surgeon and fully annotated by another expert surgeon with two types of surgical activities: phases and steps. Out of these fully annotated videos,  20 videos were selected for computational analysis.

Adverse events, according to the EAES classification were annotated too. The annotation ontology of the right hemicolectomy procedure consists of 12 phases and 6 finer-grained steps.

### 2.3 Experiments

**Preprocessing**

According to video duration, the 20 video dataset was split into training, validation, and test sets consisting of 12, 3, and 5 videos, respectively. Data characteristics of the dataset are presented in Table 2**.** On average, the surgery lasts 1 hour 47 minutes ± 31 minutes and the dataset is composed of a total number of frames at 1 fps amounts to 128,991. The images are resized to ResNet-50 input dimension of 224 x 224, and the training dataset is augmented by applying horizontal flip, saturation, and rotation.

| **Dataset** | **Videos (n)** | **Min. duration (minutes)** | **Max. duration (minutes)** | **Mean +- STD duration (minutes)** | **Total (n)** |
|---|---|---|---|---|---|
| **Training** | 12 | 52 | 144 | 101 +- 28 | 73308 |
| **Validation** | 3 | 81 | 156 | 123 +- 31 | 22165 |
| **Test** | 5 | 72 | 171 | 111 +- 34 | 33518 |

**Table 2.** Dataset statistics across the 3 data splits.

**Methodology**

A state-of-the-art deep learning model, MTMS-TCN, proposed in [14] for surgical activity recognition is utilized in this work. MTMS-TCN is a two-stage network composed of a Convolutional Neural Network (CNN) (ResNet-50) model for visual feature extraction followed by a multi-stage causal Temporal Convolutional Network (TCN) to refine the features and extracting temporal information for recognizing surgical activities.

Spatial Model: ResNet-50 [15] is one of the popular CNN architectures that has been heavily utilized in the computer vision community for activity recognition. Due to its success, ResNet-50 is utilized as a visual feature extractor and trained on images extracted from the surgical videos.

Temporal model: MTMS-TCN, is a two-stage TCN model that was trained in a multi-task learning setup on video features extracted from the CNN model. Furthermore, each stage of the TCN model consists of causal convolution that utilizes only information from past frames.

Furthermore, dilated convolutions are utilized in each layer with exponentially increasing dilation factor that facilitates capturing long temporal dependencies.

### Training Setup

The backbone ResNet-50 model is initialized with pre-trained ImageNet weights and trained for 30 epochs with a learning rate of 1e-05 and batch size of 64. Then subsequently, image features are extracted from the backbone and grouped into respective videos. The temporal model, LSTM or MTMS-TCN [16-17] is trained for the task of phase recognition on the extracted features for 200 epochs with a learning rate of 3e-04. All the models were implemented in Pytorch and trained on NVIDIA GeForce RTX 2080 Ti GPUs.

### 2.4 Critical view of safety definition and visualization (V-View)

As described for laparoscopic cholecystectomy [18], a critical view of safety can also be applied to the phase of identification and ligation of the ileocolic vessels, at the very beginning of the LRH procedure. In the following images a V-View of the ileocolic artery and vein is shown, to define its correct visualization. The V-View, as described by C. Strey et al [19] is shown in Figures 4-6.

1. Ileocolic artery and vein clearly prepared and isolated at their origin in a V-shaped manner (45°) (Figure 7.)
2. Superior mesenteric artery and/or vein visible (not always)
3. Duodenum visible below and behind (not always) along with the Toldt and Gerota fascia

**Figure 4.** Ileocolic vessels cord tractioned on the left by the assistant.

**Figure 5.** Preparation and isolation of the ileocolic artery and vein.

**Figure 6.** Preparation of the ileocolic vessels configuring a V-View**.**

## 3. RESULTS

Phases and steps were divided in 12 phases and 6 finer-grained steps according to the following scheme:

P1 - abdominal cavity exploration and preparation

P2 - ileocolic vessels identification and preparation

P3 - ileocolic vessels clipping and cutting

S4 - ileocolic vein clipping

S3 - ileocolic artery clipping

P4 – Toldt-Gerota window

P5 - mesocolon division

P6 - mesocolon division from above

P7 - last jejunal loop section

P8 - transverse colon stapling

P9 - ileo transverse colon anastomosis

S0 - firing stapler

S1 - angle point

S2 - first layer

S3 - second layer

P10 - closure of the mesenteric defect

P11 - right hemicolectomy packaging

P12 - cleaning and coagulation

### 3.1 Phase transition

Phase to phase transition was identified as shown in Figure 7.

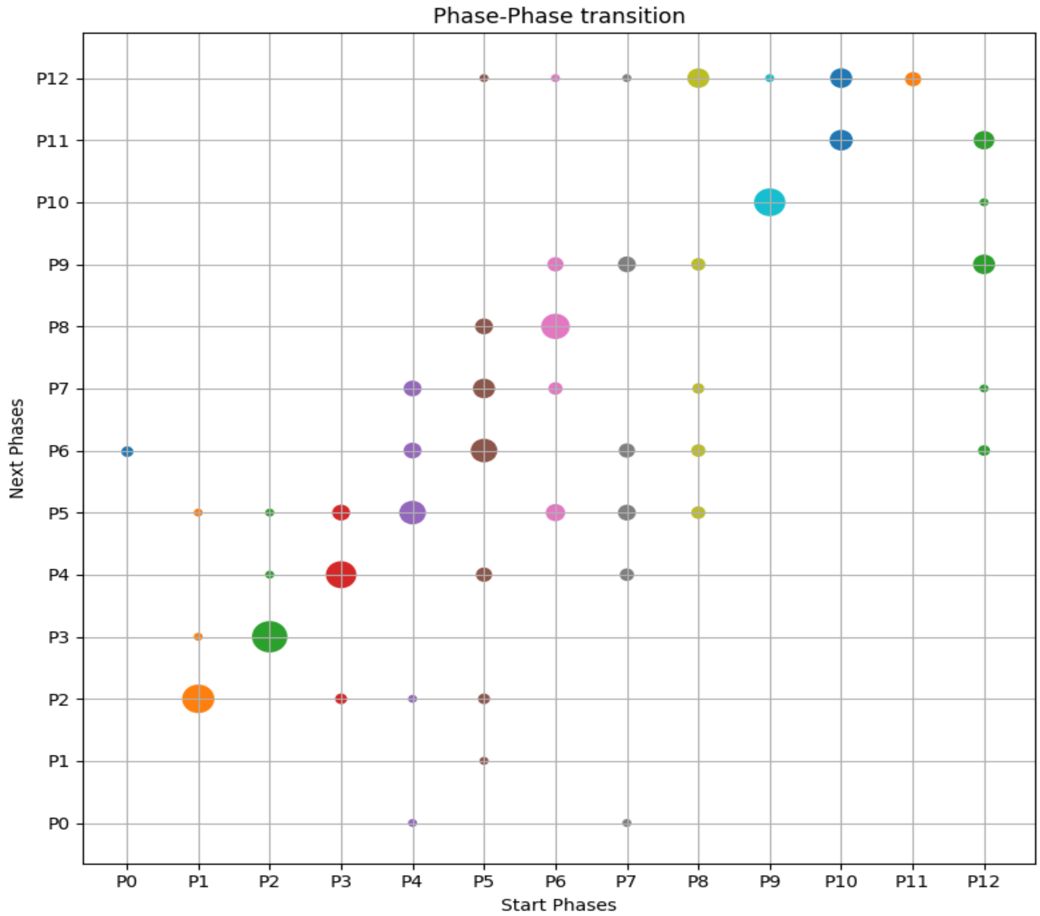


**Figure 7.** Phase-phase transition is shown in the figure above.

## 3.2 Data distribution

Phase occurrences, mean time phase duration and phase frequency are represented in Figure 8

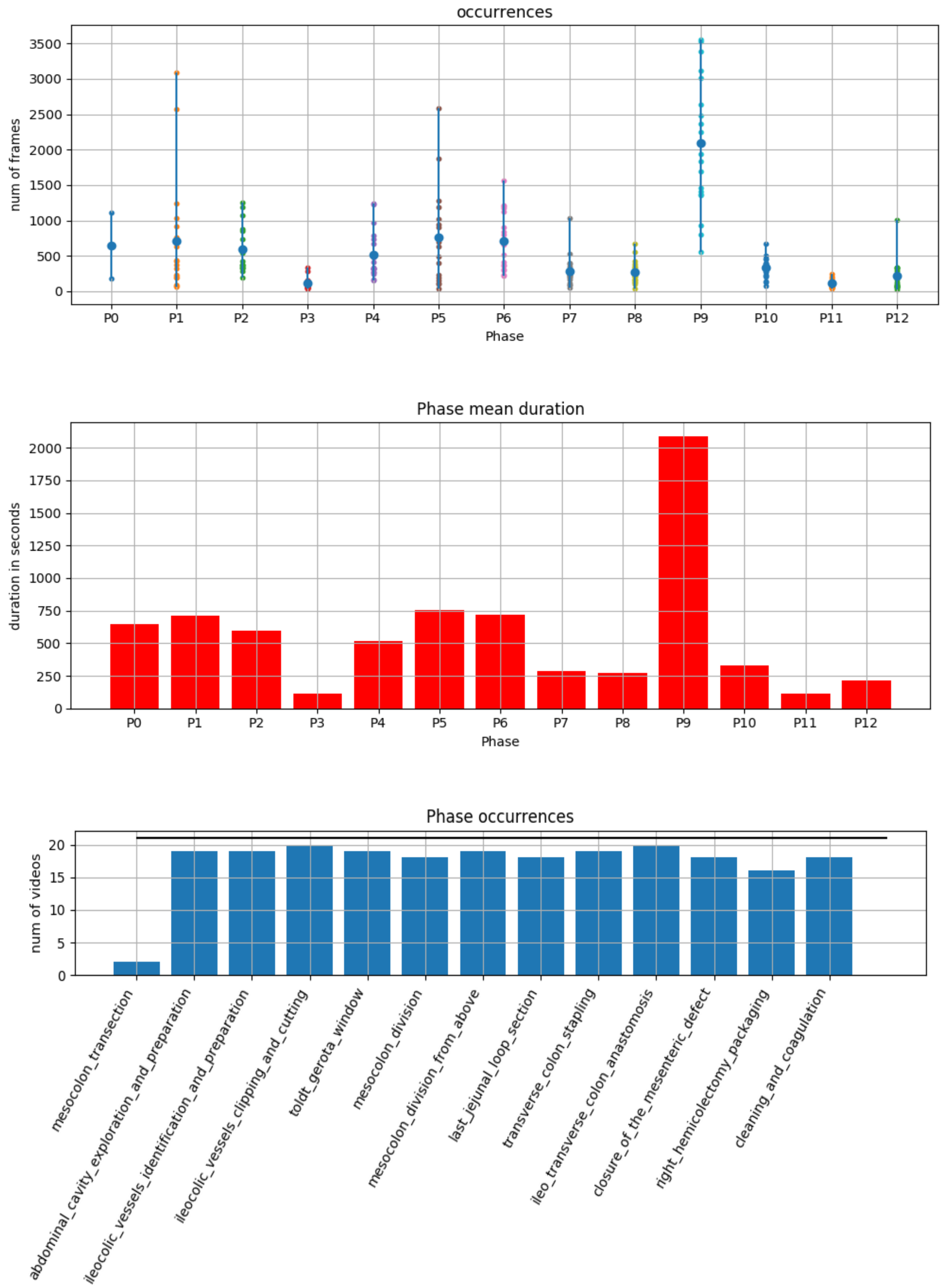


**Figure 8.** Phase occurrences, mean time phase duration and phase frequency.

### 3.3 Phase Recognition

Three models of algorithms have been trained to perform automatic phase recognition with the TCN model best performing with an accuracy of 68.25 ± 5.82 (Table 3).

| No | Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| **1** | CNN | 62.44 ± 6.05 | 45.47 ± 9.03 | 51.13 ± 8.09 | 43.84 ± 7.4 |
| **2** | LSTM | 66.74 ± 5.79 | 50.77 ± 9.78 | 61.34 ± 8.26 | 49.17 ± 8.6 |
| **3** | TCN | 68.25 ± 5.82 | 51.02 ±11.45 | 65.21 ± 9.18 | 50.59 ±10.49 |

**Table 3.** Performance in terms of accuracy and precision of the CNN, the LSTM and TCN models.

**TCN class wise**

| No | Phase | Precision | Recall | F1-score |
|---|---|---|---|---|
| 1 | P0 | 0.0 | 0.0 | 0.0 |
| 2 | P1 | 78.56 | 75.46 | 76.84 |
| 3 | P2 | 71.0 | 79.63 | 66.92 |
| 4 | P3 | 30.58 | 61.86 | 33.45 |
| 5 | P4 | 31.7 | 60.96 | 30.39 |
| 6 | P5 | 35.69 | 17.99 | 18.55 |

| | | | | |
|---|---|---|---|---|
| 7 | P6 | 54.23 | 80.23 | 57.05 |
| 8 | P7 | 35.61 | 42.56 | 33.72 |
| 9 | P8 | 27.11 | 63.98 | 34.46 |
| 10 | P9 | 94.34 | 82.76 | 87.66 |
| 11 | P10 | 79.45 | 76.89 | 74.29 |
| 12 | P11 | 66.09 | 96.52 | 77.65 |
| 13 | P12 | 34.3 | 54.81 | 38.22 |

**Table 4.** TCN model phase-specific results.

### 3.4 Phase Prediction

For a better visualization of the results in terms of accuracy in phase prediction, we used a 12 categorical color map. It can be intuitively seen that the TCN performed better than CNN in the 2 best and 2 worst videos (accuracy of 68%).

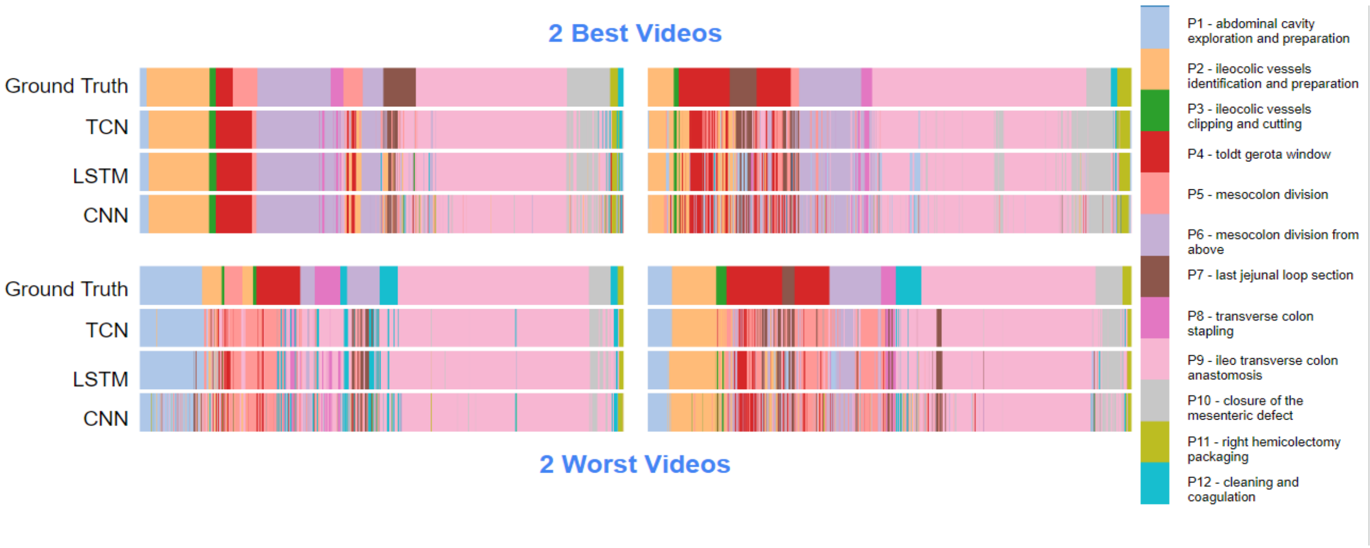


**Figure 9.** This figure visualizes a video set of 2 best and 2 worst performances of TCN-LSTM-for phase recognition.

## 4. DISCUSSION

Computer vision, the application of algorithms to analyze and interpret visual data, has become a critical technology through which to study the intraoperative phase of care with the goals of augmenting surgeons' decision-making processes, supporting safer surgery, and expanding access to surgical care. While much work has been performed on potential use cases, there are currently no CV tools widely used for diagnostic or therapeutic applications in surgery. Automated, online, laparoscopic video analysis could allow us to monitor cases in real-time, predict complications, and intervene to prevent adverse events in various fields of surgery [20].

Effective and safe surgery results from a complex sociotechnical process prone to human error. Acquiring large amounts of data on surgical care and modeling the process of surgery with artificial intelligence's computational methods could shed light on system strengths and limitations and enable computer-based smart assistance. These pioneering efforts in sensing and analyzing surgical activities, brings surgery on the verge of a fourth revolution characterized by smart assistance in perceptual, cognitive and physical tasks.

In 2020 Kitaguchi et al. first presented a large annotated dataset of laparoscopic colorectal videos with phase, action, and tool recognition with an accuracy as high as 83.2% using AI CNN model [21]. In the present study, three models of algorithms that have been trained to perform automatic phase recognition (CNN, TCN, LSTM) in LRH with the best being the TCN performing with an accuracy of $68.25 \pm 5.82$. This preliminary result suggests that research should focus on this procedure and continue adding and analyzing larger video datasets in order to improve consistency and reliability of the results. Further steps will be to extend the same process and workflow analysis to 50 videos of laparoscopic left hemicolectomy (LLH) procedure, according to a dedicated protocol.

## 4.1 Surgical applications

### Quality improvement

Postoperatively, models for procedure and surgical phase recognition could be used to automatically generate structured and segmented databases to assist with quality improvement initiatives. Such databases would represent an invaluable resource for surgical documentation. Automated video analysis could be used to digest these large collections of surgical videos, retrieve meaningful video sequences, and extract significant information. For example, full-length surgical videos can be analyzed with phase and tool detection models to identify intraoperative events and effectively produce short videos selectively documenting the division of the ileocolic vessels. Very recently, cutting-edge methods have enabled overcoming such barriers by allowing video-to-video retrieval, the task of using a video to search for videos with similar events [22]. In addition, models for phase recognition can also be used directly to automatically generate standardized surgical reports of LRH. In laparoscopic cholecystectomy cases, it has been demonstrated that incorrectly recognized video frames, i.e. model failures, could indicate complications such as bleeding or problems with gallbladder retrieval [23] CV models can be trained to extract more information from videos such as operative difficulty. Ward et al. trained a CNN to classify gallbladder inflammation according to the Parkland grading scale, a 5-tiered system based on anatomical changes. This classification then contributed to predictions of events such as bile leakage from the gallbladder during surgery and provided insights on how increases in inflammation correlate to prolonged operative times [24].

### Operative assessment and intraoperative decision support

CV models for tool detection have been used to assess the technical skills of surgeons.

In the future such models could be able to provide the surgeon with alerts in case of detected anomalies which could lead to surgical complications.

We envision the uptake of AI to assist during minimally invasive procedures. In this setting, real-time predictions from CV models could be used to guide trainees and enhance surgeon performance.

## 4.2 Limitations

Barriers to implement this vision exist and despite the considerable number of methods for automated analysis of LC videos presented in the last few years, no CV application is currently widely used in surgery.

The limiting factor for most clinical applications is the availability of well-annotated datasets. To develop effective clinical solutions, AI models are often trained to replicate expert performance from large quantities of well-annotated data. While leading to unprecedented results in medical image analysis, this learning paradigm is highly dependent on the availability of large annotated datasets [25]. Its applicability is severely limited by regulatory constraints on data-sharing and the availability of surgeons to annotate the data. These issues are further compounded by the need to well-represent and account for variations between patients, operative technique, and OR data acquisition systems.

# 5. CONCLUSIONS

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 add clinical value to CRC treatment, by eventually predicting and avoiding adverse events.

In our study the best CV algorithm was able to identify surgical phases of LRH with an accuracy of 68%. In other surgical procedures it has been demonstrated that CV could identify operative phases with accuracy similar to surgeons. Research still needs to be done to improve and develop more performing algorithms, as the computational methodologies are constantly evolving. The main challenge remains the training of such computational models which requires a relatively large number of surgeon-annotated images and high-powered computing, which are not readily available in many operating rooms globally. Subsequently, the translation of such technology to clinical setups will require an indefinite time. Ethical and legal issues should be taken into account too. Future research will consist of surgical phase, steps and adverse events, automatic recognition using larger video datasets and improved algorithm models, which will lead to potential uses in real life clinical applications such as automatic video indexing, surgical skill assessments and adverse event prediction.

## 6. BIBLIOGRAPHY

[1] T.G. Weiser et al. Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. Lancet (2015).

[2] P. Mascagni, N. Padoy. OR black box and surgical control tower: Recording and streaming data and analytics to improve surgical care. Journal of Visceral Surgery Volume 158, Issue 3, Supplement, June 2021.

[3] T. M. Ward, P. Mascagni, Y. Ban et al. Computer vision in surgery. Surgery 2021 May;169(5):1253-1256. doi: 10.1016/j.surg.2020.10.039.

[4] Bergquist S, Brooks G, Keating N, et al. Classifying lung cancer severity with ensemble machine learning in health care claims data. Proceedings of Machine Learning Research. 2017

[5] Soguero-Ruiz C, Hindberg K, Rojo-Alvarez JL, et al. Support Vector Feature Selection for Early Detection of Anastomosis Leakage From Bag-of-Words in Electronic Health Records.

IEEE J Biomed Health Inform. 2016; 20(5):1404–15.

[6] Modifi R, Duff MD, Madhavan KK, et al. Identification of severe acute pancreatitis using an artificial neural network. Surgery. 2007; 141:59–66.

[7] Monsalve-Torra A, Ruiz-Fernandez D, Marin-Alonso O, et al. Using machine learning methods for predicting inhospital mortality in patients undergoing open repair of abdominal aortic aneurysm. J Biomed Inform. 2016; 62:195–201.

CAMBIARE DA QUI

[8] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017; 542(7639):115–118.

[9] Soguero-Ruiz C, Hindberg K, Mora-Jimenez I, et al. Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods. J Biomed Inform. 2016; 61:87–96.

[10] Balvardi, S. et al. The association between video-based assessment of intraoperative technical performance and patient outcomes: a systematic review. Surg. Endosc. 10.1007/s00464-022-09296-6 (2022).

[11] Feldman LS, et al. SAGES Video-Based Assessment (VBA) program: a vision for life-long learning for surgeons. Surg. Endosc. 2020;34: 3285–3288.

[12] EAES classification of intraoperative adverse events in laparoscopic surgery. Francis NK, Curtis NJ, Conti JA, Foster JD, Bonjer HJ, Hanna GB; EAES committees. Surg Endosc. 2018 Sep;32(9):3822-3829. doi: 10.1007/s00464-018-6108-1.

[13] Nathan J Curtis, Godwin Dennison, Chris S B Brown et al. Clinical Evaluation of Intraoperative Near Misses in Laparoscopic Rectal Cancer Surgery. Ann Surg 2021 Apr 1;273(4):778-784.3.

[14] Ramesh, S., Dall'Alba, D, Gonzalez et al. Multi-task temporal convolutional networks for joint recognition of surgical phases and steps in gastric bypass procedures. International Journal of Computer Assisted Radiology and Surgery (2021).

[15] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2016).

[16] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. In Neural Computation (Vol. 9, Issue 8, pp. 1735–1780). MIT Press - Journals.

[17] Sanat Ramesh, Diego Dall'Alba, Cristians Gonzalez et al. TRandAugment: temporal random augmentation strategy for surgical activity recognition from videos. Int J Comput Assist Radiol Surg 2023 Mar 22.

[18] Mascagni P, Alapatt D, Laracca GG, Guerriero L. Multicentric validation of EndoDigest: a computer vision platform for video documentation of the critical view of safety in laparoscopic cholecystectomy. SN. Surg Endosc. 2022 Nov;36(11):8379-8386. 336

[19] C. W. Strey, C. Wullstein, M. Adamina et al. Laparoscopic right hemicolectomy with CME: standardization using the "critical view" concept. Surg Endosc. 2018 Dec;32(12):5021-5030.

[20] Yu, T. & Padoy, N. Encode the Unseen: Predictive Video Hashing for Scalable Mid-stream Retrieval. In: Ishikawa, H., Liu, CL., Pajdla, T., Shi, J. (eds) Computer Vision – ACCV 2020. ACCV 2020. Lecture Notes in Computer Science, vol 12626. Springer, Cham. 10.1007/978-3-030-69541-5_26 (2021).

[21] Kitaguchi D, Takeshita N, Matsuzaki H, Oda T, Watanabe M, Mori K, et al. Automated Laparoscopic Colorectal Surgery Workflow Recognition using Artificial Intelligence: Experimental Research. Int J Surg 2020. https://doi.org/10.1016/j.ijsu.2020.05.015.

[22] Yu, T. et al. Live laparoscopic video retrieval with compressed uncertainty. Preprint at:

https://arxiv.org/abs/2203.04301 (2022).

[23] Berlet, M. et al. Surgical reporting for laparoscopic cholecystectomy based on phase annotation by a convolutional neural network (CNN) and the phenomenon of phase flickering: a proof of concept. Int. J. Comput. Assist. Radiol. Surg. 10.1007/s11548-022-02680-6 (2022)

[24] Ward, T. M., Hashimoto, D. A., Ban et al. O. R. Artificial intelligence prediction of cholecystectomy operative course from automated identification of gallbladder inflammation. Surg. Endosc. 10.1007/s00464-022-09009-z (2022).

[25] Esteva A, et al. A guide to deep learning in healthcare. Nat. Med. 2019; 25:24–29