Process Mining in Surgical Workflow Analysis from Videos

FRANCISCO THOMAS VAN DER BOON, University of Twente, The Netherlands

Each year, there are more than 200 million surgeries performed worldwide, providing essential and often lifesaving treatment. This study explores the application of process mining techniques to improve the analysis of surgical workflows using an annotated laparoscopic dataset. The dataset includes three types of colorectal surgeries: Proctocolectomy, Rectal Resection and Sigmoid Resection.

Process mining, which derives workflow models from event logs, was applied using heuristic mining and alignment-based conformance checking to construct standardized surgical process models. These models serve as empirical baselines, enabling comparison of individual procedures against data-driven best-practice workflows.

The analysis revealed recurring misalignments in specific surgical phases and substantial variability in phase transitions and procedure durations. these findings highlight the potential of process mining to quantitatively evaluate surgical performance, informing training programs, and supporting outcome monitoring.

Additional Key Words and Phrases: Process Mining • Colorectal Surgery • Workflows • Laparoscopic • Heuristic Miners

1 INTRODUCTION

Process mining is a relatively recent subfield of data science, introduced in the late 1990's by prof.dr.ir. Wil van der Aalst [1]. It revolves around the analysis of event logs, a structured record of activities within a process, to extract and construct process models. The analysis of these models consists of three main stages: *process discovery*, *conformance checking* and *process enhancement* [17]. Process discovery refers to the creation of process models based on observed event logs, making use of various algorithms known as miners. These models are then evaluated, validated and analysed using conformance checking, which quantifies the differences between the models and new event logs. Finally, process enhancement refers to the improvement or extension of the models using new data, with the goal of increasing the alignments of the model with the real-world behaviour. Through this framework, process mining allows for the transformation of raw event data into quantifiable and visual process representations.

In the domain of healthcare, process mining has been increasingly prominent, providing optimizations for clinical procedures. Prior studies have used it for case studies, process optimizations and analysis and most notably in the context of our research, workflow analysis [13]. By recording and visualising procedural events (traces), process mining can offer a means to better understand the variability in surgeries and the various outcomes resulting from that variability.

Each year, there are more than 200 million surgeries performed worldwide, providing essential and often lifesaving treatment [7]. Although modern surgical techniques are generally safe, complications and procedural errors are remain a concern, posing significant challenge across surgical contexts [7]. Establishing a standardized workflow derived from real surgical data offers a means for surgeons to analyse risks of upcoming procedures and improve surgical outcomes. By aggregating different procedures across multiple instances of the same surgical type, it becomes possible to construct an accurate representative model of a colorectal surgery. These models can provide guidance during training and help identify critical deviations and phases that can influence patient outcomes.

Author's address: Francisco Thomas van der Boon, f.t.vandervanderboon@student.utwente.nl, University of Twente, P.O. Box 217, Enschede, The Netherlands, 7500AE.

TScIT 43, July 4, 2025, Enschede, The Netherlands

2 PROBLEM STATEMENT

There is currently no universally established standard for surgical procedures, particularly in the context of teaching and guidelines [7]. This lack of standardization can result in variability across surgical practices, which can lead to avoidable risks before, during and after the procedure. In complex procedures, such as colorectal surgery, defined workflows can be a valuable asset for ensuring consistency and improving outcomes.

Colorectal surgical complications can generally be categorised into intraoperative complications and postoperative complications [8]. One of the most critical postoperative complications is anastomotic leakage, linked to an increased mortality rate [9]. Research has also demonstrated a positive correlation between increased operating time and both intraoperative and postoperative complications [8]. While intraoperative complications are linked to increased operating times [9], they are also a predictor for poorer surgical outcomes. Interestingly, studies have shown that prolonged operating times in laparoscopic surgeries have not contributed heavily to the increased intraoperative and postoperative complications, in contrast to the alternative - open surgeries, likely due to the inherent advantages that laparoscopic surgery provides [8]. Additionally, evidence suggests a learning curve for laparoscopic colorectal surgeries exists, with operating times and intraoperative complications declining steadily after approximately 30 procedures [15].

2.1 Research Questions

This research inversitages how surgical workflow models from video data can be improved with the additional help of tenchniques such as process mining. By constructing standardized reference models, we can compare individual procedures to an empirical baseline, and quantify the variance. This analysis may contribute to better understanding of surgical procedures and consequences, leading to increased training and guidelines, potentially resulting in the reduction of procedural complications.

(1) How can surgical workflows be improved with the help of techniques from process mining?
  (a) How can process mining be a viable asset in surgical workflow mining from computer vision algorithms?
  (b) To what extent can alignment-based conformance checking reveal patterns in surgical performance variability?

3 TECHNICAL BACKGROUND

This research aims to analyse the application and effectiveness of process mining in the context of surgical workflows. Specifically, the study explores whether process mining algorithms can be meaningfully applied to a colorectal dataset derived from a computer vision model.

3.1 Process Mining

Process mining, introduced in the late 20th century by prof.dr.ir. Wil van der Aalst, consists of three general stages: process discovery, conformance checking and process enhancement [17]. In the process discovery stage, event logs are analysed to generate a process model that is representative of the real-world procedure that was provided. A key challenge is to determine and select an appropriate mining algorithm (miner) to derive the most meaningful process model in accordance with the dataset. Conformance checking then evaluates and validates the process model by comparing it to a different event log of the same procedure, providing a quantifiable measure of model accuracy. It also highlights the main differences between the model and event log, providing key insight into deviations from expected and observed behaviours. Finally, process enhancement uses the alignment diagnostics from the previous stage to improve or extend the model, ensuring it accurately describes the real-world process [17].

This study focused primarily on the process discovery and conformance checking stages of process mining. The central consideration during process discovery was selecting a suitable mining algorithm that would best model the procedure's behaviour. Three types of miners were evaluated: Integer Linear Programming (ILP) miner, the inductive miner, and the heuristic miner.

The ILP miner aims to find a model that perfectly fits the event log that is provided as input, often resulting in a fitness of 1. This is considered a drawback by most, however for this research, where avoiding deviations from the real world in the model is the goal, it was a suitable option. In our case, the resulting ILP model incorrectly permitted access to the beginning and end of the procedure at any phase, it further introduced an unrealistic flexibility to the miner along with an alignment later that favoured the model over the event log.

Subsequently, we examined the inductive miner, which recursively splits the event log to identify patterns [18]. While this approach can offer improved generalisation and accuracy with more data, it's structure becomes overly complex in a sparse dataset. Finally, the heuristic miner was also considered for this research, the key difference between the heuristic miner and the inductive miner is the dependency graph that is generated before creating the splits and identifying the patterns [4]. After comparative testing, we found that this difference resulted in the heuristic miner producing a slightly better F-score and weighted average than the inductive miner, and was therefore selected as our model for the next phase of process mining.

3.2 Computer Vision

This annotated dataset in this study reflects the type of output that can be provided by a Computer Vision (CV) algorithm applied to surgical videos. A specific challenge for surgical phase labelling is detecting the start of a phase. The dataset follows three definitions that defined the start of a phase [12]:

- A phase starts when the instrument related to the first activity relevant for this phase enters the screen.
- If a change of the anatomical region results in the transition to a new phase, the camera movement towards the region marks the start of the phase.
- If the camera leaves the body or is pulled back into the trocar between two phases, the new phase starts with the first frame that does not show the trocar in which the camera is located.

These criteria imply that both surgical instruments and camera location influence phase labelling. The Robust Medical Instrument Segmentation (ROBUST-MIS) challenge was proposed to develop an international benchmark for robust and generalized algorithms to detect and categorize surgical instruments [14]. As the instrument segmentation and detection is available, the location of the camera would be the next step to improve reliability of automated surgical phase recognition.

3.3 Related Work

Process mining in health care has seen it's fair share of research, surgical workflows are not an exception. In this section, we are going over some of the related work for process mining in surgical workflows.

We are using an open source dataset from the Heidelberg University Hospital [11]. The dataset has a collection of four different surgeries all focused around colorectal surgeries. The dataset consists of frame, phase label pairs, where each of the 13 phase labels represent the highest level of surgical workflow analysis [12].

Christos Spiliadis, a Master student at the Technical University of Delft, wrote a master thesis on surgical workflow analysis [16]. The research was conducting by using ceiling mounted cameras followed by a combination of GMM-HMM models, to automate the phase creation of surgical workflow.

Web-video-mining-supported workflow modelling, is a paper published in 2016 focusing on surgical workflow modelling from web-video mining [10]. They created a scraping algorithm and combined it with a selection and segmentation algorithm that allowed them to create a probabilistic surgical workflow model.

Modelling the workflow of a cholecystectomy has also been used to build a Hidden Markov Model [3]. The research resulted in a GUI that graphically represented the surgical workflow, allowing the comparison of surgical video that have been synchronized with the model.

4 EXPERIMENTAL METHODOLOGY

The methodology was structured into three semi-sequential phases: pre-processing, process discovery, and conformance checking. This modular design enabled a clear structure, and backtracking was only required when a new miner needed to be tested or further pre-processing was implemented. The overall workflow is illustrated in Figure 1, which depicts the components and their interdependencies.

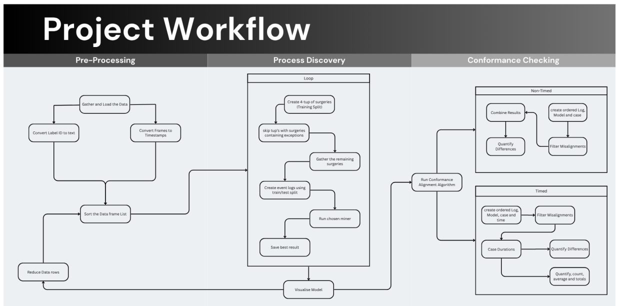

Fig. 1. Project methodology flow chart

4.1 Pre-Processing

The pre-processing phase began with a review of the dataset to identify the existing elements and how they can be transformed into the required structure for event log construction. Event logs are composed of three core attributes: a case identifier, an activity label, and a timestamp. The data we received contains two distinct columns: *Frames* and *Labels*. To this end, the procedure identifier (e.g., *Proctocolectomy_1_Phase*) was used as the case ID. The 13 unique phase labels were mapped to their distinct textual representation to improve interpretability and reduce confusion. The frame number was converted into a timestamp, spanning a 1 second : 25 frames ratio, using the current implementation time as a starting point.

Each procedure contained, on average, 270'000 rows. Most of these entries are redundant, as each phase can range between seconds and several minutes. To reduce this duplication, for both memory efficiency and to improve the clarity of model construction, only the first instance and the time of the last instance of a set of consecutive phases was retained. This preserved the full temporal scope of the procedure, while reducing the retention of duplication. The transformation condensed the event logs to fewer than 100 rows per procedure, preserving all relevant information.

4.2 Process Discovery

The process discovery was central to the research, producing the reference model required later for conformance checking as well as providing visual insight into surgical procedures and patterns. Three primary goals defined this step: (1) identifying a suitable mining algorithm, (2) constructing a suitable reference model, and (3) creating a train/test split consisting of a set of 10 procedures of the same type.

Initially the experimentation focused on selecting an appropriate miner. Based on literature, the ILP miner appeared as a promising candidate due to its well-known characteristic of producing models with a perfect fitness (fitness = 1.0). This initially suggested that it would closely follow the structure of the event log, creating a reference model perfectly mimicking the real world. However, empirical testing revealed the opposite. The ILP miner over generalized the model, introducing two "black boxes", which reduced the interpretability and precision of the model. Through the black boxes, each of the phases was allowed to start and end the entire procedure. It further resulted in a model focused alignment result making the ILP miner unviable for our application.

Subsequent testing introduced both the inductive and the heuristic miners. The heuristic miner outperformed the inductive miner, on their respective best working models, in terms of F-score and weighted average metrics. The F-score was computed using the standard harmonic mean between fitness and precision [6]:

$$F\_score = \frac{2 * fitness * precision}{fitness + precision}$$

To further refine the model selection, a weighted average of the four process mining quality metrics, fitness (f), precision (p), generalization (g), and simplicity (s), was constructed using the following formula [2]:

$$WA = 0.4 * f + 0.4 * p + 0.15 * g + 0.05 * s$$

The weights reflect our priority of balancing model accuracy with interpretability and generalization.

To automate the generation of the train/test split, a script was implemented to automatically generate and evaluate all valid combinations. A ratio of 4:6 was chosen, to prioritize the quantification and analysis of data. To ensure that the generated model reflected a typical surgical procedure, all combinations containing a procedure with the phase 'exception' were skipped. The algorithm returned the train/test split with the highest average between F-score and weighted average, corresponding to the selected reference model reflected in this study.

4.3 Alignment-based Conformance Checking

The final methodological phase involved evaluating the individual surgical traces and how they aligned with the discovered reference model. This was accomplished by making use of *pm4py.conformance_diagnostics_alignments* module. This algorithm identifies an optimal alignment between a process model and an event log by minimizing the cost of deviations, thereby producing a trace-by-trace mapping of observed vs. expected behaviour. Due to the characteristics of the heuristic miner, the reference model contained many black boxes, resulting in an alignment that favoured log moves as more cost efficient, reproducing an alignment that contained the full event log.

Each event log results in a sequence of tuples in the form *(log move, model move)*, representing how each trace aligns or misaligns with a corresponding step in the process model. Custom scripts were written for this type of output to generate dataframes containing log moves, model moves and the associated case identifier, which served as the basis for the analysis and visualisations.

Due to the nature of this dataset, conformance checking was conducted in two forms: once focusing on **non-timed results** and once with a focus on **timed results**. This enabled separate evaluations of structural alignments and temporal deviation and anomalies, without cluttering either result with too much information.

A critical aspect of the alignment output was the identification and filtering of misalignments. To highlight the deviations, two types of tuples were excluded:

- **Synchronous moves**, where the model and the event log transitions aligned, indicating expected behaviour.
- **Artificial start transitions**, returned as a *(">>", None)* tuple, which typically occurs and the beginning of a trace representing a "skip" on both sides and is therefore not meaningful.

Upon removing these elements, the resulting data structure retains only the deviations:

- **Log moves**, which are traces that are observed but not expected by the model.
- **Model moves**, which are traces that are expected by the model but not observed in the event log.

These filtered misalignments help highlight and identify patterns of structural deviations and potential anomalies in the procedures.

5 EXPERIMENTAL SETUP

5.1 Dataset

The study used the 'Heidelberg Colorectal Dataset for surgical data Science in the sensor operating room' (HeiCo) [11, 12, 14]. This dataset covers three types of colorectal procedures: Proctocolectomy, Rectal Resection and Sigmoid Resection, each having 10 separate recordings. This has resulted in 30 fully annotated laparoscopic colorectal procedures with an average duration of approximately 3 hours. The HeiCo dataset provided a structured and semantically rich foundation, similar to a computer vision algorithm, from which valid event logs could be constructed for process mining.

5.2 General Approach

The first milestone was constructing a representative reference model of one of the surgery types. While there are certain established conventions for surgical procedures, most executions vary in practice due to patient-specific factors and intraoperative decisions and complications. Consequently the reference model constructed followed a generalized sequence derived from the observed data, preserving a real world structure of events.

To validate the robustness and representativeness of the reference model, a train/test split was employed. All combinations of training and testing splits were evaluated systematically, through the script, each resulting in a model scored using the F-score [5] and weighted average metrics.

After selecting the optimal model based on the validation metrics, the next milestone is the alignment-based conformance checking. This step enables granular comparison between the observed event log and the expected reference model, at the level of individual event transitions.

Lastly, post-processing allowed the transformation of the results into analysable and interpretable data structures.

5.3 Implementation

The research was carried out in a Jupyter notebook environment, using python 3.11.x and pm4py 2.7.x. The main requirement to recreate the result is a large CPU, as process mining requires a lot of memory to run efficiently.

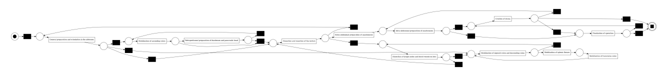

Fig. 2. Best heuristic result

6 RESULTS

This section presents the finding of the alignment-based conformance checking and explores two parallel analyses: **non-timed** and **timed** alignment results. The non-timed results highlight phase sequences and misalignment patterns, while the timed results allow exploration of phase durations, repetitions and temporal deviations across surgical procedures. These same analyses were also performed on the full alignment to gain further insight into the differences and similarities of the full procedures.

6.1 Reference Model

The reference model, constructed using the heuristic miner on procedures 1, 4, 6, and 7, is a Petri-net visualized in Figure 2. The model includes several transition (black box) nodes, highlighting the flexibility in these surgical procedures. This model allows for varied paths between phases, creating closed loops and defined sequences.

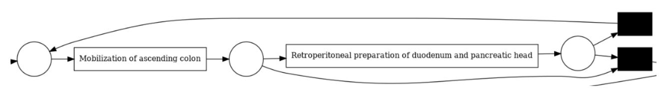

Fig. 3. Phase 7 to phase 3 transition in heuristic workflow model

An area of interest involves the interaction between phase 7 and phase 3 (Figure 3). Clinically, these phases are often linked, as phase 3 may serve as an intermediate step during the mobilization of the ascending colon (Online Table 1) [12]. The model, showing transitions from phase 7 to both phase 3 and a black box, reflects this variability. This suggests that phase 3, which only has an incoming connection from phase 7, is fully interconnected with that phase.

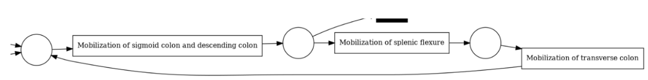

Fig. 4. Phase 4, 5, 6 transition in heuristic workflow model

A second noteworthy region of the model concerns the typical sequence form phase 4 to phase 5, and onward to phase 6 (Figure 4). The model reflects a consistent pattern observed across all training procedures, with transitions confirming the expected phase progression. A similar pattern was also discovered between phase 9 and phase 10 (Figure 5), representing the two most aligned phases according to our results.

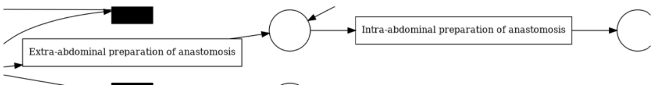

Fig. 5. Phase 9 to phase 10 transition in heuristic workflow model

Table 1. Durations of each surgery

Procedure	Duration (HH-MM-SS)
1	04-07-44
2	03-53-41
3	03-05-58
4	02-54-21
5	03-31-57
6	02-16-04
6	02-50-21
8	02-53-08
9	03-31-55
10	04-56-21

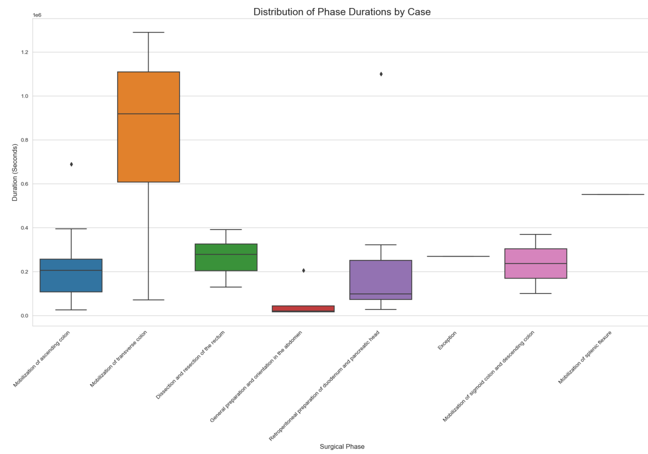

Fig. 6. Misaligned Combined Bar Chart

6.2 Non-Timed Alignment Results

The non-timed conformance results provide insights into misalignments based purely on phase transitions, irrespective of their respective durations. Two procedures, 3 and 10, display prominent deviations from the reference model.

Procedure 10 is the longest in Proctocolectomy dataset, lasting nearly 5 hours. Its alignment showed a significant deviation during the middle of the operation. Specifically the sequence:

$$Phase\_7 \rightarrow Phase\_0 \rightarrow Phase\_3$$

was represented multiple times, despite being an invalid transition in the reference model. No exception was recorded for this case, suggesting an operational challenge. The repetition of phase 0 implies potential difficulties completing or progressing beyond specific steps, further suggesting operational challenges, likely stemming from patient conditions.

Procedure 3 highlights a similar anomaly. An exception, though unspecified of the exact nature, occurred midway through phase 6. This could be the explanation as to why the surgical team addressed phase 6 before phase 7, a further deviation from our model. While the nature of the exception is unknown, the likely outcome was an adaption in the procedure not known to our reference model.

6.3 Timed Alignment Results

To complement the structural analysis, a timed evaluation of the results was performed. Table 1 presents the total durations for each Proctocolectomy procedure. Timed analysis enables the identification of phase variability and commonality.

6.3.1 Total and Average phase durations. Phase 8 emerged as the most complex and variable across procedures. It was repeated an average of three times per surgery and each exhibiting durations ranging from 12 to 45 minutes. The high variance, both in total and average duration, supports the hypothesis that this phase is technically demanding and heavily dependent on patient-specific anatomy.

Procedure 10 exhibited significant misalignment inside of which, phase repetitions were highly notable: six repetitions for two phases and three for others. One instance of phase 6 lasted over an hour, while combined with the remaining deviations the phases accounted for over two hours of the surgery, nearly the full duration of an average Proctocolectomy procedure. This pattern reinforces earlier observations that this procedure represents a significant outlier in this surgery type. It further suggests, that during this procedure there was likely a patient-specific anatomy that induced a high degree of complexity to the procedure.

6.3.2 Phase Variability and Stability. Figure 6 presents an overview of misaligned trace durations across phases and cases. The most volatile phase was again found to be phase 6, particularly influenced by procedure 10, demonstrated by Figure 10. Looking deeper into the misalignments, a difference between the one-minute and 17-minute separation was found. This suggests that at least one instance may have been mislabelled, due to the annotation conditions, potentially due to brief spatial proximity without active engagement. This hypothesis is further supported by Figures 11 and 12, which both highlight the one-minute phase as an outlier specific to procedure 10.

In contrast, several phases appear relatively stable, both in the misaligned and full-trace analyses. These phases often only occur once per surgery, hinting at procedural simplicity or anatomical constraints. Phase 7, for example, while frequently involved in misalignment transitions, remained consistent in duration, suggesting a simple yet frequently revisited step. Similarly, other than the outlier in procedure 10, phase 3 can be observed as a highly stable phase, reinforcing the hypothesis above, as the non-timed results showed phase 7 to be a wrapper for phase 3.

A broader observation arises when comparing misaligned versus full trace. Phase 8, which was found to be one of the most complex and variable phases, had few yet very stable misalignments. Compared to phase 6, which we found to have very high variability, when looking at the full surgical procedure, it was found to be much more stable compared to phase 8. This indicates that model alignment, while useful in identifying patterns, does not necessarily correlate with surgical difficulty. Instead, variability in full procedure traces may reflect more complex challenges, reinforcing the need for both structural and temporal perspectives in workflow evaluation.

7 DISCUSSION

7.1 Interpretation of key findings

This study set out to explore how surgical workflows can be improved through the application of process mining techniques. Section 4 outlines the methodology used to achieve this, demonstrating how annotated video data can be transformed into structured event logs and process models. The results presented in section 6 show that even with a small dataset, process mining revealed both typical procedural sequences and meaningful deviations. Patterns such as the misalignment in procedure 10 and the high variability in phase 8 illustrate how process mining can support structured workflow evaluation by identifying variations and points of interest within surgical practice.

Two sub-questions supported this central inquiry. The first sub-question asked whether process mining could be combined with CV algorithms to become a viable asset in surgical workflow mining. While CV was not implemented in this study, the structure of the HeiCo dataset and the resulting models suggest that integration and automation is feasible. If developed further, CV integration could enable scalable, near real-time workflow analysis, reducing reliance on manual annotation and traditional statistical models.

The second sub-question addressed the extent to which alignment-based conformance checking could quantify surgical variability. The results indicate that this technique is effective in visualizing differences across procedures and within phases. When paired with clinical context such as patient characteristics or surgical outcomes, these deviations can support surgeon reflection, performance evaluation and quality improvement efforts.

Overall, these findings suggest that process mining is a promising tool for surgical workflow analysis. With future integration of CV, it could support real-time feedback, surgical training, and continuous quality assurance by enabling automated and timely evaluation of surgical procedures.

7.2 Limitations

This study has faced several limitations related to the scope and structure of the dataset, which constrained the depth of analysis. Notably, phase 13, in two of the three types of surgery was only observed in a single procedure. As a result, it was not feasible to investigate the potential causes or implications of these exceptions in a meaningful or statistically grounded way.

A fundamental characteristic of process mining is the availability of a large number of cases with relatively fewer activities. This structure allows algorithms to detect robust patterns and dependencies with greater accuracy. However, in this study, the opposite was true: the dataset contained more activity types than distinct surgical procedures. The most detailed procedure only consists of 36 meaningful rows after pre-processing. Consequently, the resulting models were based on sparse data, and the connections derived from them remain largely hypothetical rather than conclusive.

Furthermore, the dataset lacked contextual metadata, such as patient demographics, or surgical outcomes. This omission limited our ability to interpret the causes of prolonged phases and misalignments. Without this information, it is not possible to differentiate between variations due to procedural complications and those arising from patient-specific factors. Thus, while the observed patterns and deviations may be accurate reflections of surgical behaviour, their definitive explanation remains speculation.

7.3 Future Work

There are several promising directions for future work. First and foremost, expanding the dataset, as mentioned in the limitations, would significantly enhance the value of the results and analysis, allowing for less speculative answers. Moreover, with appropriate implementation, introducing this to surgical teams could serve as valuable education and reflection on the procedures that are done. Ultimately, integrating this approach into surgical data science pipelines, with the use of CV, could contribute to improved patient outcomes by identifying inconsistencies or opportunities for best-practice reinforcement.

8 CONCLUSION

This study demonstrated that process mining techniques can be applied to laparoscopic video-derived data to generate and analyse surgical workflows. Even with a limited dataset, the resulting models captured both standard procedural structure and variation. Alignment-based conformance checking revealed patterns of surgical variability, offering potential value for training and performance evaluations. With further development, particularly the integration of computer vision for more efficient annotation, process mining could become a valuable component of surgical data science, supporting real-time assessment and continuous improvement.

REFERENCES

[1] [n. d.]. Process Mining. https://www.processmining.org/introduction.html

[2] Jan Niklas Adams and Wil M.P. Van Der Aalst. 2021. Precision and Fitness in Object-Centric Process Mining. In *2021 3rd International Conference on Process Mining (ICPM)*. IEEE, Eindhoven, Netherlands, 128–135. https://doi.org/10.1109/ICPM53251.2021.9576886

[3] Tobias Blum, Nicolas Padoy, Hubertus Feußner, and Nassir Navab. 2008. Workflow mining for visualization and analysis of surgeries. *International Journal of Computer Assisted Radiology and Surgery* 3, 5 (Nov. 2008), 379–386. https://doi.org/10.1007/s11548-008-0239-0

[4] Sofie De Cnudde, Jan Claes, and Geert Poels. 2014. Improving the quality of the Heuristics Miner in ProM 6.2. *Expert Systems with Applications* 41, 17 (Dec. 2014), 7678–7690. https://doi.org/10.1016/j.eswa.2014.05.055

[5] Jochen De Weerdt, Manu De Backer, Jan Vanthienen, and Bart Baesens. 2011. A robust F-measure for evaluating discovered process models. In *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. IEEE, Paris, France, 148–155. https://doi.org/10.1109/CIDM.2011.5949428

[6] Leon Derczynski. 2016. Complementarity, F-score, and NLP Evaluation, Vol. Proceedings of the Tenth International Conference on Language Resources and Evaluation ({LREC}'16). opean Language Resources Association (ELRA), Portorož, Slovenia, 261–266. https://aclanthology.org/L16-1040/

[7] Fernando J. Kim, Rodrigo Donalisio Da Silva, Diedra Gustafson, Leticia Nogueira, Timothy Harlin, and David L. Paul. 2015. Current issues in patient safety in surgery: a review. *Patient Safety in Surgery* 9, 1 (Dec. 2015), 26. https://doi.org/10.1186/s13037-015-0067-4

[8] Philipp Kirchhoff, Pierre-Alain Clavien, and Dieter Hahnloser. 2010. Complications in colorectal surgery: risk factors and preventive strategies. *Patient Safety in Surgery* 4, 1 (Dec. 2010), 5. https://doi.org/10.1186/1754-9493-4-5

[9] Magdalena A. Lipska, Ian P. Bissett, Bryan R. Parry, and Arend E. H. Merrie. 2006. ANASTOMOTIC LEAKAGE AFTER LOWER GASTROINTESTINAL ANASTOMOSIS: MEN ARE AT A HIGHER RISK. *ANZ Journal of Surgery* 76, 7 (July 2006), 579–585. https://doi.org/10.1111/j.1445-2197.2006.03780.x

[10] Rui Liu, Xiaoli Zhang, and Hao Zhang. 2016. Web-video-mining-supported workflow modeling for laparoscopic surgeries. *Artificial Intelligence in Medicine* 74 (Nov. 2016), 9–20. https://doi.org/10.1016/j.artmed.2016.11.002

[11] Lena Maier-Hein, Martin Wagner, Tobias Ross, Annika Reinke, Sebastian Bodenstedt, Peter M. Full, Hellena Hempe, Diana Mindroc-Filimon, Patrick Scholz, Thuy Nuong Tran, Pierangela Bruno, Anna Kisilenko, Benjamin Müller, Tornike Davitashvili, Manuela Capek, Minu D. Tizabi, Matthias Eisenmann, Tim J. Adler, Janek Gröhl, Melanie Schellenberg, Silvia Seidlitz, T. Y. Emmy Lai, Bünyamin Pekdemir, Veith Roethlingshoefer, Fabian Both, Sebastian Bittel, Marc Mengler, Lars Mündermann, Martin Apitz, Annette Kopp-Schneider, Stefanie Speidel, Felix Nickel, Pascal Probst, Hannes G. Kenngott, and Beat P. Müller-Stich. 2020. Heidelberg Colorectal Data Set for Surgical Data Science in the Sensor Operating Room. https://doi.org/10.7303/SYN21903917

[12] Lena Maier-Hein, Martin Wagner, Tobias Ross, Annika Reinke, Sebastian Bodenstedt, Peter M. Full, Hellena Hempe, Diana Mindroc-Filimon, Patrick Scholz, Thuy Nuong Tran, Pierangela Bruno, Anna Kisilenko, Benjamin Müller, Tornike Davitashvili, Manuela Capek, Minu D. Tizabi, Matthias Eisenmann, Tim J. Adler, Janek Gröhl, Melanie Schellenberg, Silvia Seidlitz, T. Y. Emmy Lai, Bünyamin Pekdemir, Veith Roethlingshoefer, Fabian Both, Sebastian Bittel, Marc Mengler, Lars Mündermann, Martin Apitz, Annette Kopp-Schneider, Stefanie Speidel, Felix Nickel, Pascal Probst, Hannes G. Kenngott, and Beat P. Müller-Stich. 2021. Heidelberg colorectal data set for surgical data science in the sensor operating room. *Scientific Data* 8, 1 (April 2021), 101. https://doi.org/10.1038/s41597-021-00882-2

[13] Eric Rojas, Jorge Munoz-Gama, Marcos Sepúlveda, and Daniel Capurro. 2016. Process mining in healthcare: A literature review. *Journal of Biomedical Informatics* 61 (June 2016), 224–236. https://doi.org/10.1016/j.jbi.2016.04.007

[14] Tobias Roß, Annika Reinke, Peter M. Full, Martin Wagner, Hannes Kenngott, Martin Apitz, Hellena Hempe, Diana Mindroc-Filimon, Patrick Scholz, Thuy Nuong Tran, Pierangela Bruno, Pablo Arbeláez, Gui-Bin Bian, Sebastian Bodenstedt, Jon Lindström Bolmgren, Laura Bravo-Sánchez, Hua-Bin Chen, Cristina González, Dong Guo, Pål Halvorsen, Pheng-Ann Heng, Enes Hosgor, Zeng-Guang Hou, Fabian Isensee, Debesh Jha, Tingting Jiang, Yueming Jin, Kadir Kirtac, Sabrina Kletz, Stefan Leger, Zhixuan Li, Klaus H. Maier-Hein, Zhen-Liang Ni, Michael A. Riegler, Klaus Schoeffmann, Ruohua Shi, Stefanie Speidel, Michael Stenzel, Isabell Twick, Gutai Wang, Jiacheng Wang, Liansheng Wang, Lu Wang, Yujie Zhang, Yan-Jie Zhou, Lei Zhu, Manuel Wiesenfarth, Annette Kopp-Schneider, Beat P. Müller-Stich, and Lena Maier-Hein. 2021. Comparative validation of multi-instance instrument segmentation in endoscopy: Results of the ROBUST-MIS 2019 challenge. *Medical Image Analysis* 70 (May 2021), 101920. https://doi.org/10.1016/j.media.2020.101920

[15] Christopher M. Schlachta, Joseph Mamazza, Pieter A. Seshadri, Margherita Cadeddu, Roger Gregoire, and Eric C. Poulin. 2001. Defining a learning curve for laparoscopic colorectal resections. *Diseases of the Colon & Rectum* 44, 2 (Feb. 2001), 217–222. https://doi.org/10.1007/BF02234296

[16] C. Spiliadis. 2025. *Surgical Workflow Analysis An Explainable Approach.* Master's thesis. TU Delft. https://resolver.tudelft.nl/uuid:a402564a-0d58-4bcd-9a45-5f186b2d3e0e

[17] Wil Van Der Aalst. 2012. Process Mining: Overview and Opportunities. *ACM Transactions on Management Information Systems* 3, 2 (July 2012), 1–17. https://doi.org/10.1145/2229156.2229157

[18] Eric Varbeek and Joos Buijs. [n. d.]. Inductive Miner. https://www.futurelearn.com/info/courses/process-mining/0/steps/15642

A TABLES

A.1 Phase Label

Table 2. Phase ID to textual label

ID	Phase name
0	General preparation and orientation in the abdomen
1	Dissection of lymph nodes and blood vessels en bloc
2	Retroperitoneal preparation towards lower pancreatic border
3	Retroperitoneal preparation of duodenum and pancreatic head
4	Mobilization of sigmoid colon and descending colon
5	Mobilization of splenic flexure
6	Mobilization of transverse colon
7	Mobilization of ascending colon
8	Dissection and resection of the rectum
9	Extra-abdominal preparation of anastomosis
10	Intra-abdominal preparation of anastomosis
11	Creation of stoma
12	Finalization of operation
13	Exception

A.2 Non-times results

Table 3. Misaligned traces

2_Phase	3_Phase	5_Phase	8_Phase	9_Phase	10_Phase	Log Move Count	Unique Phases	Normalized Difference
7	None	None	None	None	None	1	1	0
3	None	None	8	None	None	2	2	0
7	None	8	None	8	None	3	2	0.5
None	6	4	8	4	None	4	3	0.333333333
6	13	None	4	None	None	3	3	0
5	7	7	None	7	None	4	2	1
None	3	3	None	3	None	3	1	2
None	7	7	None	7	7	4	1	3
None	3	3	None	3	0	4	2	1
None	8	7	7	7	7	5	2	1.5
None	None	3	3	3	0	4	2	1
None	None	8	None	8	7	3	2	0.5
None	None	None	None	None	0	1	1	0
None	None	None	None	None	3	1	1	0
None	None	None	None	None	6	1	1	0
None	None	None	None	None	0	1	1	0
None	None	None	None	None	0	1	1	0
None	None	None	None	None	6	1	1	0

Table 4. Full event log alignment trace

2_Phase	3_Phase	5_Phase	8_Phase	9_Phase	10_Phase	Log Move Count	Unique Phases	Normalized Difference
0	0	0	0	0	0	6	1	5
7	None	None	None	None	None	1	1	0
3	4	4	8	4	4	6	3	1
7	5	8	4	8	5	6	4	0.5
None	6	4	8	4	6	5	3	0.666666667
6	13	5	4	5	None	5	4	0.25
4	6	6	5	6	5	6	3	1
5	None	None	None	None	6	2	2	0
6	None	None	4	None	None	2	2	0
5	7	7	5	7	5	6	2	2
4	3	3	6	3	6	6	3	1
None	7	7	None	7	7	4	1	3
None	3	3	None	3	0	4	2	1
None	8	7	7	7	7	5	2	1.5
8	1	3	3	3	0	6	4	0.5
None	None	8	1	8	7	4	3	0.333333333
0	8	1	None	1	0	5	3	0.666666667
None	9	None	8	None	3	3	3	0
8	10	8	9	8	4	6	4	0.5
9	None	9	10	9	None	4	2	1
10	11	10	None	10	6	5	3	0.666666667
None	None	None	11	None	0	2	2	0
12	10	11	None	11	6	5	4	0.25
None	None	None	10	None	0	2	2	0
11	12	12	None	12	6	5	3	0.666666667
None	None	None	12	None	4	2	2	0
None	None	None	None	11	None	1	1	0
None	None	None	11	None	1	2	2	0
None	None	None	None	12	None	1	1	0
None	None	None	None	None	8	1	1	0
None	None	None	None	None	9	1	1	0
None	None	None	None	None	10	1	1	0
None	None	None	None	None	11	1	1	0
None	None	None	None	None	12	1	1	0

A.3 Timed results

case_id	phase	activity_count	average_duration	total_duration
1_Phase	0	2	04:56.5	00:09:53
1_Phase	1	3	02:24.2	07:12.6
1_Phase	3	1	13:15.3	13:15.3
1_Phase	4	4	05:44.1	22:56.6
1_Phase	5	1	01:57.3	01:57.3
1_Phase	6	2	13:38.1	27:16.2
1_Phase	7	2	05:40.1	11:20.1
1_Phase	8	5	12:14.3	01:11.6
1_Phase	9	1	52:05.4	52:05.4
1_Phase	10	1	19:44.1	19:44.1
1_Phase	11	1	19:22.3	19:22.3
1_Phase	12	1	01:28.6	01:28.6

10_Phase	0	6	02:08.9	12:53.3
10_Phase	1	1	05:12.8	05:12.8
10_Phase	3	1	18:19.8	18:19.8
10_Phase	4	3	11:47.0	35:21.0
10_Phase	5	3	05:04.8	15:14.4
10_Phase	6	6	09:34.9	57:29.6
10_Phase	7	3	05:22.0	16:05.9
10_Phase	8	1	45:20.7	45:20.7
10_Phase	9	1	43:21.4	43:21.4
10_Phase	10	1	30:08.3	30:08.3
10_Phase	11	1	14:44.6	14:44.6
10_Phase	12	1	02:07.7	02:07.7
2_Phase	0	2	02:49.1	05:38.1
2_Phase	3	1	04:50.0	04:50.0
2_Phase	4	2	14:28.2	28:56.4
2_Phase	5	2	06:30.6	13:01.3
2_Phase	6	2	13:56.3	27:52.6
2_Phase	7	2	07:29.4	14:58.8
2_Phase	8	2	31:21.2	02:42.4
2_Phase	9	1	53:12.4	53:12.4
2_Phase	10	1	10:26.9	10:26.9
2_Phase	11	1	02:30.1	02:30.1
2_Phase	12	1	09:31.1	09:31.1
3_Phase	0	1	02:29.9	02:29.9
3_Phase	1	1	10:28.3	10:28.3
3_Phase	3	2	02:59.0	05:58.0
3_Phase	4	1	09:36.8	09:36.8
3_Phase	5	1	06:19.2	06:19.2
3_Phase	6	2	08:19.3	16:38.6
3_Phase	7	2	02:46.1	05:32.3
3_Phase	8	2	22:32.9	45:05.7
3_Phase	9	1	38:06.4	38:06.4
3_Phase	10	2	13:00.7	26:01.4
3_Phase	11	1	07:24.1	07:24.1
3_Phase	12	1	07:47.7	07:47.7
3_Phase	13	1	00:04:29	00:04:29
4_Phase	0	3	02:40.4	08:01.3
4_Phase	1	2	00:03:23	00:06:46
4_Phase	3	2	01:33.7	03:07.4
4_Phase	4	1	12:28.7	12:28.7
4_Phase	5	1	00:09:42	00:09:42
4_Phase	6	1	11:50.5	11:50.5
4_Phase	7	2	02:43.3	05:26.6
4_Phase	8	3	12:19.7	36:59.2
4_Phase	9	1	39:03.0	39:03.0
4_Phase	10	2	15:29.4	30:58.8
4_Phase	11	2	02:36.3	05:12.7
4_Phase	12	2	02:21.8	04:43.6
5_Phase	0	1	08:32.4	08:32.4
5_Phase	1	1	06:57.5	06:57.5
5_Phase	3	3	00:55.9	02:47.8
5_Phase	4	2	04:16.8	08:33.6
5_Phase	5	1	03:16.0	03:16.0

5_Phase	6	1	16:23.8	16:23.8
5_Phase	7	3	01:58.3	05:54.8
5_Phase	8	3	19:33.2	58:39.6
5_Phase	9	1	47:26.8	47:26.8
5_Phase	10	1	22:24.8	22:24.8
5_Phase	11	1	14:32.1	14:32.1
5_Phase	12	1	16:27.1	16:27.1
6_Phase	0	1	00:59.4	00:59.4
6_Phase	1	2	01:48.8	03:37.6
6_Phase	3	1	02:13.4	02:13.4
6_Phase	4	1	06:56.8	06:56.8
6_Phase	5	1	01:35.4	01:35.4
6_Phase	6	1	09:03.8	09:03.8
6_Phase	7	2	02:55.3	05:50.6
6_Phase	8	3	15:58.6	47:55.9
6_Phase	9	1	32:30.9	32:30.9
6_Phase	10	2	06:55.5	13:50.9
6_Phase	11	2	04:10.3	08:20.5
6_Phase	12	2	01:33.9	03:07.8
7_Phase	0	2	04:53.0	09:46.0
7_Phase	1	1	07:10.6	07:10.6
7_Phase	3	2	01:59.2	03:58.5
7_Phase	4	1	09:04.5	09:04.5
7_Phase	5	1	03:32.6	03:32.6
7_Phase	6	1	14:35.2	14:35.2
7_Phase	7	3	00:58.7	02:56.1
7_Phase	8	3	12:05.4	36:16.3
7_Phase	9	1	41:01.5	41:01.5
7_Phase	10	2	12:03.3	24:06.6
7_Phase	11	2	07:28.0	14:56.1
7_Phase	12	1	02:57.3	02:57.3
8_Phase	0	1	06:55.7	06:55.7
8_Phase	1	1	04:00.6	04:00.6
8_Phase	3	1	05:22.9	05:22.9
8_Phase	4	3	04:16.0	12:48.1
8_Phase	5	2	05:27.9	10:55.8
8_Phase	6	1	08:04.4	08:04.4
8_Phase	7	1	04:23.7	04:23.7
8_Phase	8	3	16:01.8	48:05.4
8_Phase	9	1	38:47.6	38:47.6
8_Phase	10	2	10:07.1	20:14.1
8_Phase	11	2	05:44.3	11:28.6
8_Phase	12	1	02:00.6	02:00.6
9_Phase	0	1	08:30.6	08:30.6
9_Phase	1	1	06:57.4	06:57.4
9_Phase	3	3	01:20.1	04:00.4
9_Phase	4	2	05:19.8	10:39.7
9_Phase	5	1	02:32.3	02:32.3
9_Phase	6	1	00:15:23	00:15:23
9_Phase	7	3	01:24.6	04:13.8
9_Phase	8	3	19:35.5	58:46.6
9_Phase	9	1	47:26.8	47:26.8
9_Phase	10	1	22:24.4	22:24.4

9_Phase	11	2	09:41.5	19:23.0
9_Phase	12	2	05:48.3	11:36.6

B ENHANCED FIGURES

B.1 Petri-net

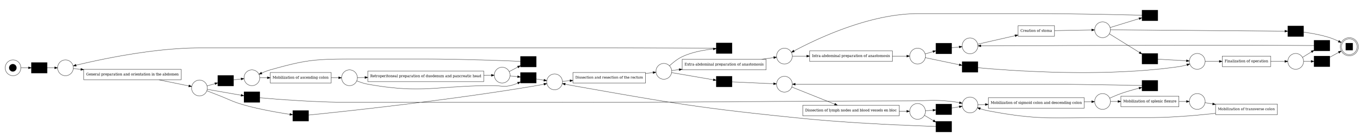

Fig. 7. Full-size version of best heuristic result

B.2 Full Sized Workflow Chart

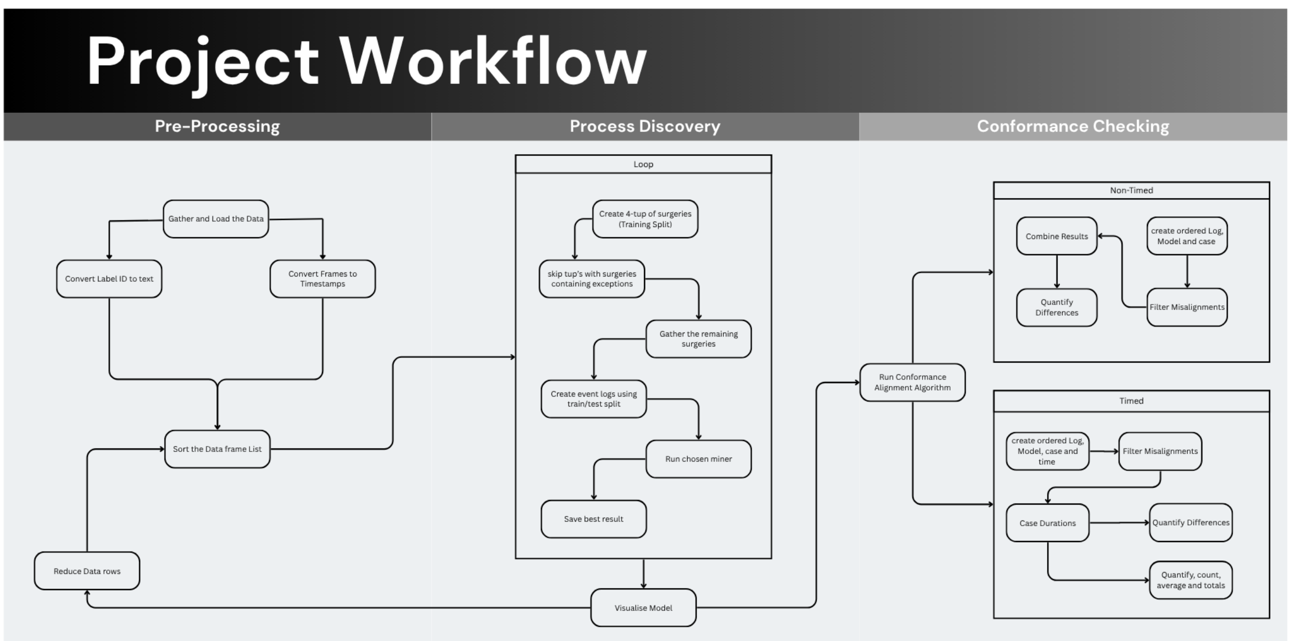

Fig. 8. Large workflow chart

B.3 Full Sized Bar Charts

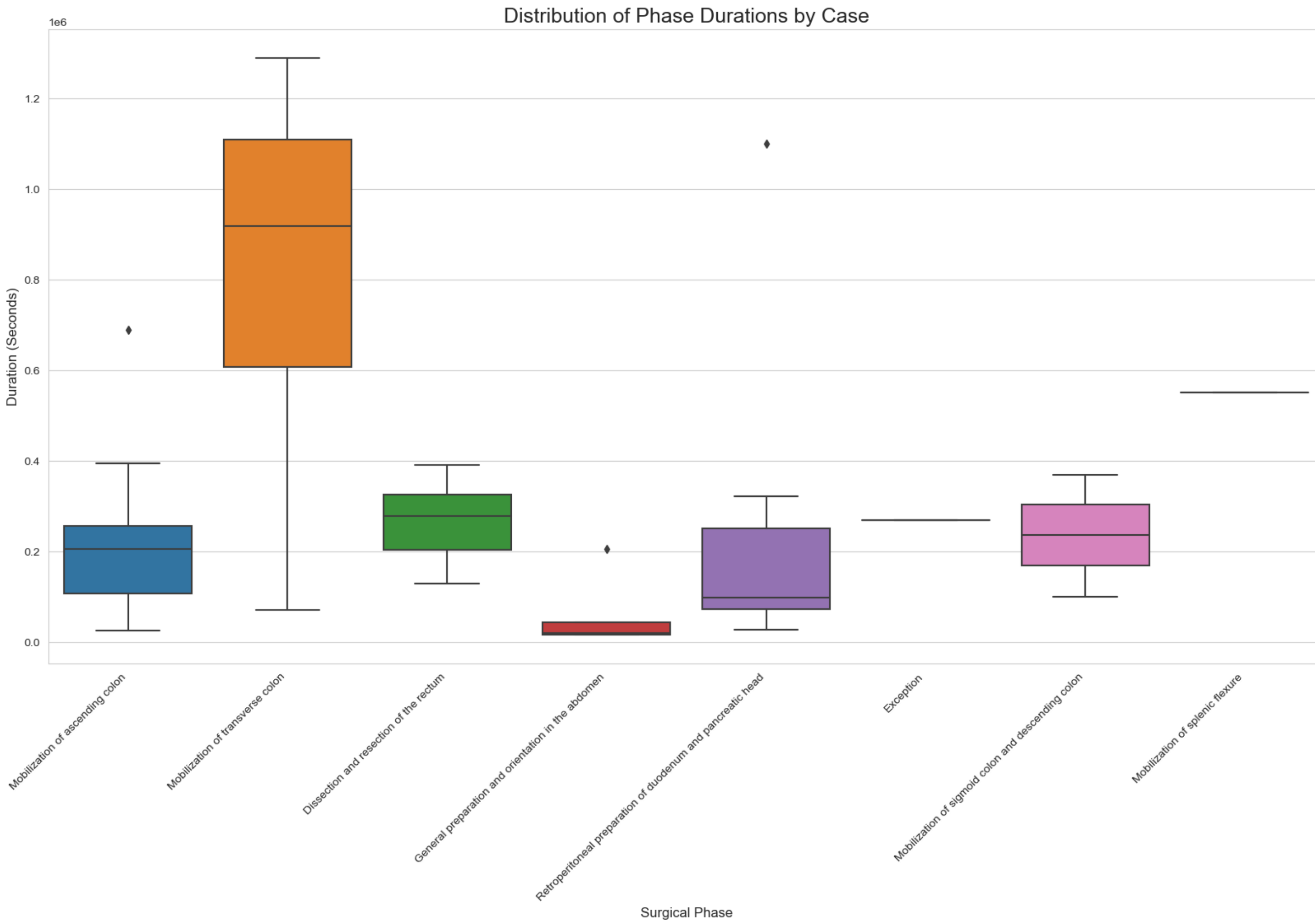

Fig. 9. Full Sized - Misaligned combined bar chart

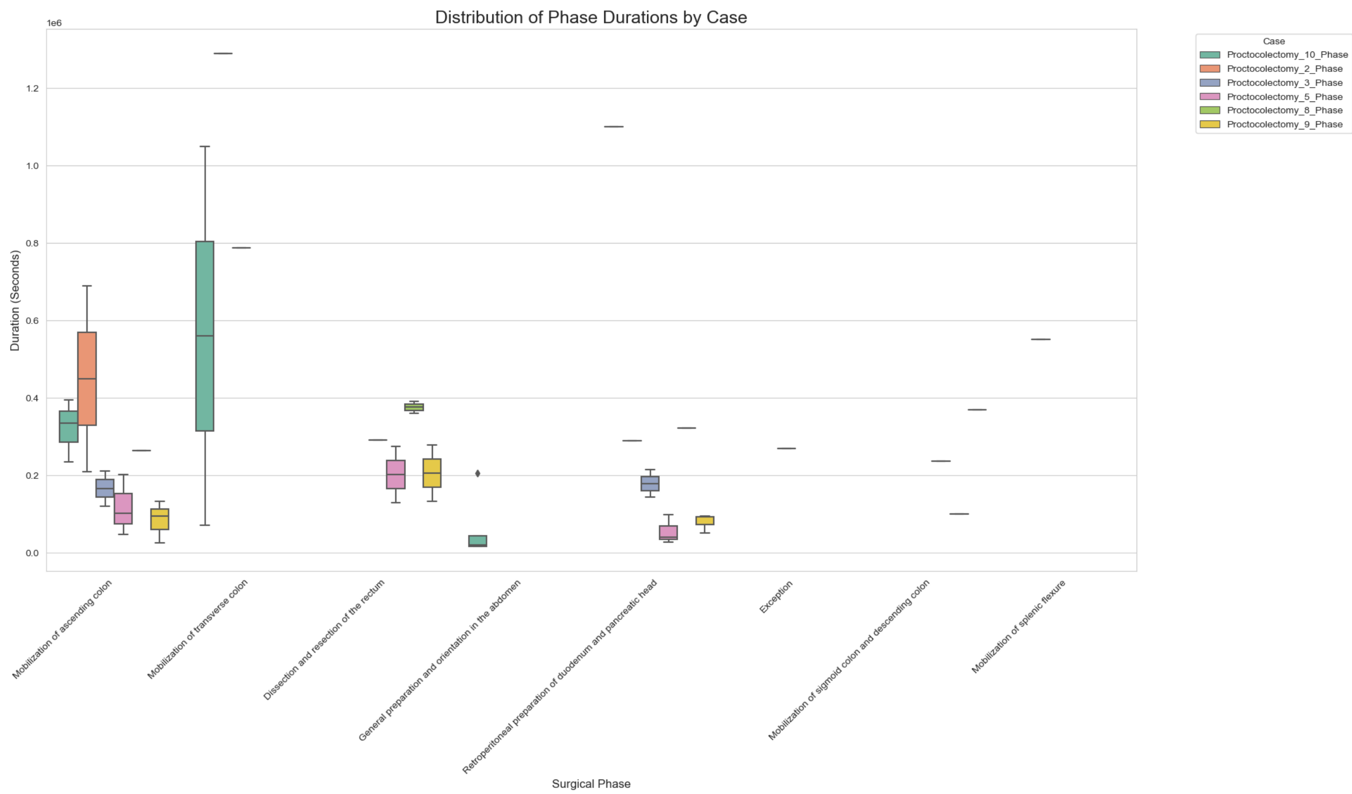

Fig. 10. Full Sized - Misaligned separated bar chart

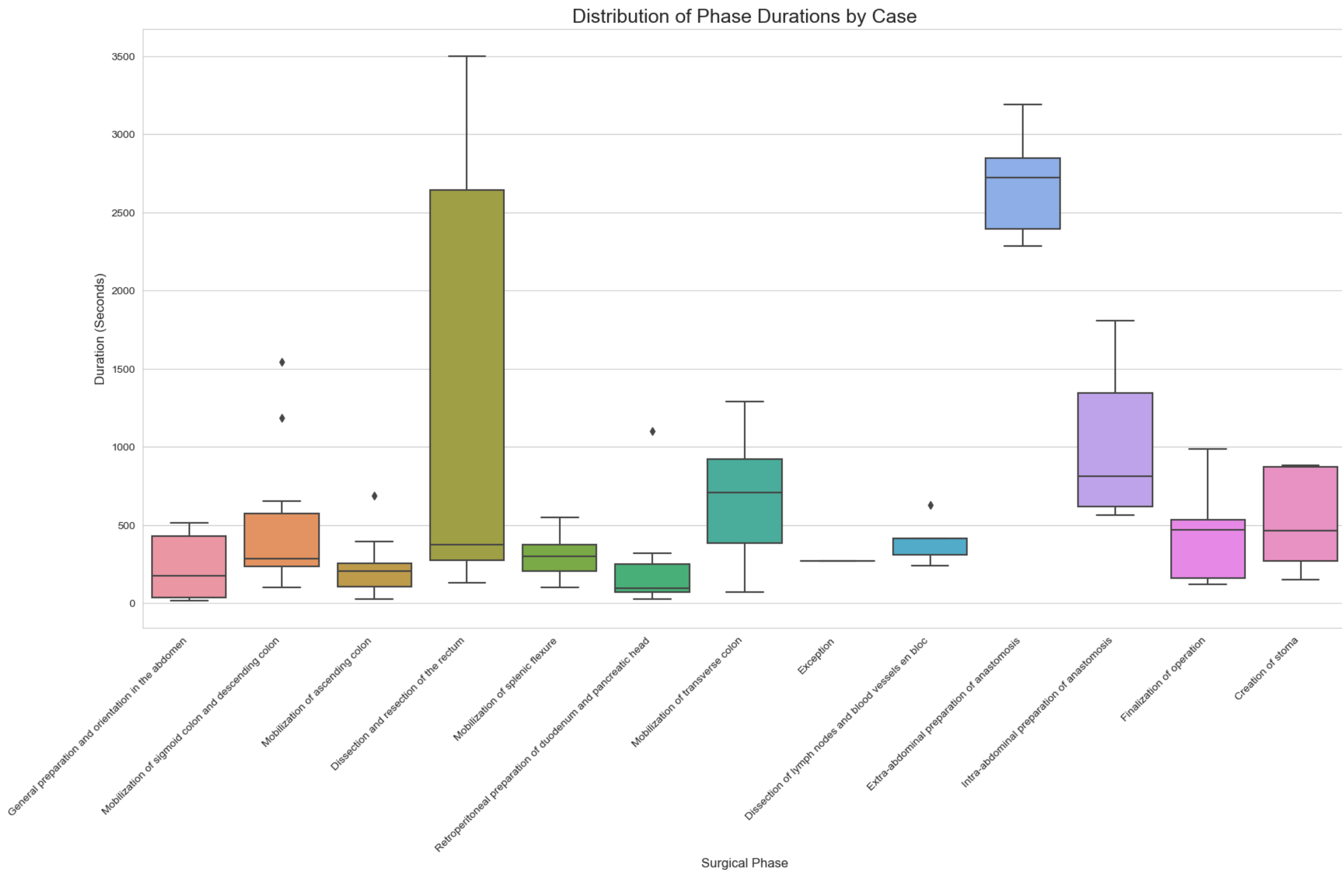

Fig. 11. Full Sized - Combined bar chart

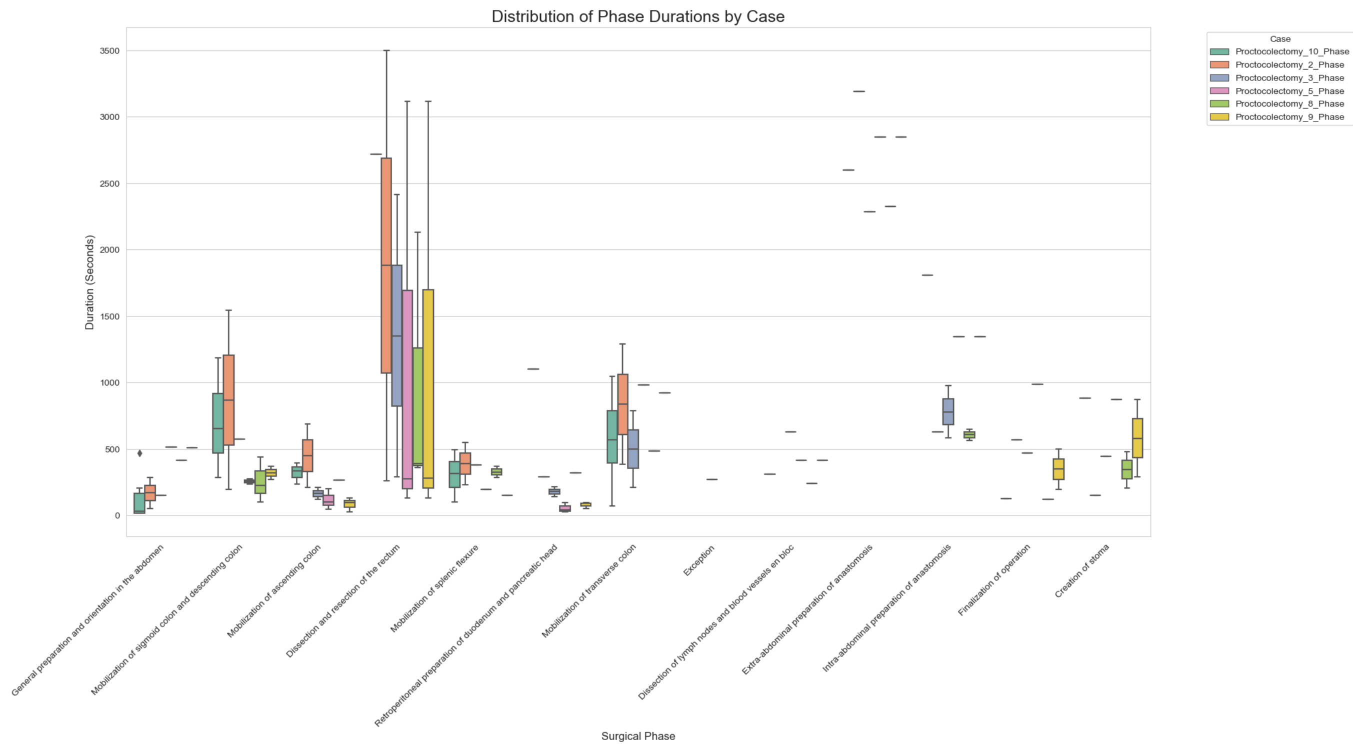

Fig. 12. Full Sized - Separated bar chart

C STATEMENT ON AI USE

During the development of this thesis, I used OpenAI's ChatGPT to support the research process in the following ways:

- To brainstorm and refine ideas related to the methodology.
- To review the paper for grammar and academic tone.

At all times, the content and analysis presented in this work reflect my own understanding, interpretation, and original contributions. The AI tool was used solely to enhance the quality and readability of the final text.