Surgical Endoscopy
https://doi.org/10.1007/s00464-022-09009-z

2021 SAGES ORAL

Artificial intelligence prediction of cholecystectomy operative course from automated identification of gallbladder inflammation

Thomas M. Ward[1] · Daniel A. Hashimoto[1] · Yutong Ban[1,2] · Guy Rosman[1,2] · Ozanan R. Meireles[1]

Received: 29 August 2021 / Accepted: 3 January 2022

Abstract

Background Operative courses of laparoscopic cholecystectomies vary widely due to differing pathologies. Efforts to assess intra-operative difficulty include the Parkland grading scale (PGS), which scores inflammation from the initial view of the gallbladder on a 1–5 scale. We investigated the impact of PGS on intra-operative outcomes, including laparoscopic duration, attainment of the critical view of safety (CVS), and gallbladder injury. We additionally trained an artificial intelligence (AI) model to identify PGS.

Methods One surgeon labeled surgical phases, PGS, CVS attainment, and gallbladder injury in 200 cholecystectomy videos. We used multilevel Bayesian regression models to analyze the PGS's effect on intra-operative outcomes. We trained AI models to identify PGS from an initial view of the gallbladder and compared model performance to annotations by a second surgeon.

Results Slightly inflamed gallbladders (PGS-2) minimally increased duration, adding 2.7 [95% compatibility interval (CI) 0.3–7.0] minutes to an operation. This contrasted with maximally inflamed gallbladders (PGS-5), where on average 16.9 (95% CI 4.4–33.9) minutes were added, with 31.3 (95% CI 8.0–67.5) minutes added for the most affected surgeon. Inadvertent gallbladder injury occurred in 25% of cases, with a minimal increase in gallbladder injury observed with added inflammation. However, up to a 28% (95% CI − 2, 63) increase in probability of a gallbladder hole during PGS-5 cases was observed for some surgeons. Inflammation had no substantial effect on whether or not a surgeon attained the CVS. An AI model could reliably (Krippendorff's $\alpha = 0.71$, 95% CI 0.65–0.77) quantify inflammation when compared to a second surgeon ($\alpha = 0.82$, 95% CI 0.75–0.87).

Conclusions An AI model can identify the degree of gallbladder inflammation, which is predictive of cholecystectomy intra-operative course. This automated assessment could be useful for operating room workflow optimization and for targeted per-surgeon and per-resident feedback to accelerate acquisition of operative skills.

This manuscript was a Best Papers podium presentation at the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) 2021 Annual Scientific Meeting, Las Vegas, NV, USA, which occurred from August 31–September 3, 2021.

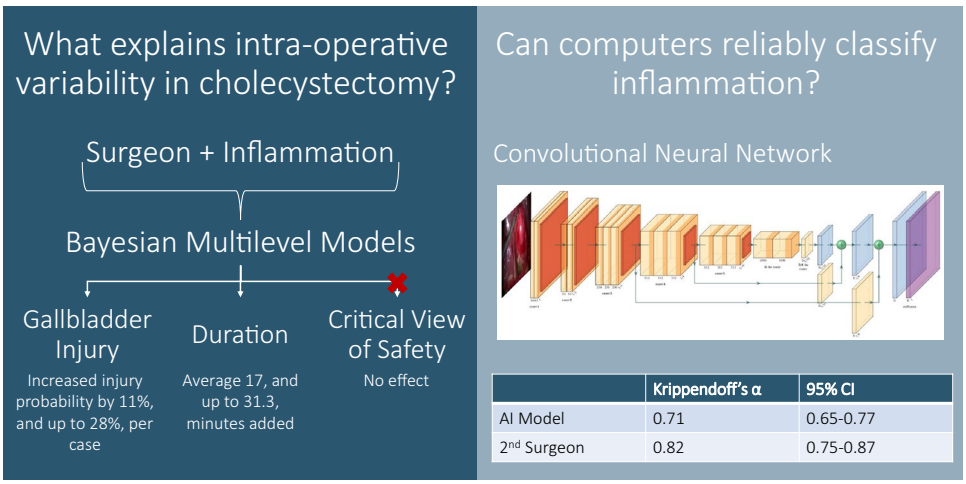
✉ Thomas M. Ward
tmward@mgh.harvard.edu

[1] Surgical Artificial Intelligence and Innovation Laboratory, Department of Surgery, Massachusetts General Hospital, 15 Parkman St., WAC 460, Boston, MA 02114, USA

[2] Distributed Robotics Laboratory, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

Published online: 14 January 2022

Graphical abstract



Keywords Computer Vision · Deep learning · Artificial intelligence · Cholecystectomy

Operative courses of laparoscopic cholecystectomies vary widely due to differing pathologies. These pathologies, which can range from symptomatic biliary colic to gangrenous cholecystitis, create unique intra-operative challenges through their differing degrees of inflammation. Many intra-operative grading scales exist to assess gallbladder inflammation, such as the G10 gallbladder scoring system, Nassar operative difficulty scale, and the Parkland grading scale (PGS) [1–3]. For example, the PGS ranks gallbladder inflammation from 1 to 5, with 5 being the worst, based on the initial intra-operative view of the gallbladder. Previous studies have investigated these scales' associations with post-operative outcomes, such as bile leak and readmission, and intra-operative outcomes, such as conversion to open surgery, operative duration, bile spillage, and bleeding [4–6]. However, these previous studies did not account for the most important driver of differences in these outcomes: the surgeon [7]. They also did not investigate if inflammation had an effect on whether or not the surgeon attained the critical view of safety (CVS).

We hypothesized that gallbladder inflammation affects the operative course. In particular, we investigated the effect of gallbladder inflammation, as quantified by the PGS, on operative duration, creation of inadvertent holes in the gallbladder, and attainment of the CVS. PGS was used, rather than the Nassar and G10 grading scales, due to its objective subcomponents and proven inter-rater reliability [6]. We used multilevel Bayesian models to allow for differing effects by surgeon and to precisely quantify the change in outcome between levels of gallbladder inflammation, as scored with the PGS. We additionally trained an Artificial Intelligence (AI) model to identify PGS to allow for real-time prediction of operative events. This AI model used computer vision, a subfield of AI that teaches machines visual comprehension.

Materials and methods

Institutional approval

This study's protocol was reviewed and approved by the Mass General Brigham Institutional Review Board (Protocol No: 2018P001641). Written patient consent for use of the videos for research purposes was obtained prior to any procedures being performed.

Dataset

We collected 200 laparoscopic cholecystectomy videos from the Massachusetts General Hospital (Boston, MA). Videos were processed and de-identified with the FFmpeg software [8]. Videos were linked to the surgeon who performed the case with an anonymized identifier. One surgeon annotated each video for operative phases, level of inflammation (PGS), attainment of the CVS, and gallbladder injury during removal from the liver bed. A second surgeon also reviewed and annotated every representative frame with a PGS rating to be used for evaluation of the computer vision models' performance. The surgical phases annotated included Port Placement, Fundus Retraction, Release of Gallbladder

Peritoneum, Dissection of Calot's Triangle, Intra-operative Cholangiogram, Cystic Artery Clipping, Cystic Artery Division, Cystic Duct Clipping, Cystic Duct Division, Removal of Gallbladder from Liver Bed, and Bagging, following guidelines defined in our prior work [9]. From the timestamps and duration of each surgical phase, additional metrics for each video were calculated, including time until first clip application, dissection duration, and laparoscopic duration (total video duration with any intra-operative cholangiogram time subtracted). The timestamp when the CVS was first completely achieved, as defined by Strasberg et al. was recorded for each video [10]. A frame from each video was extracted that best showed the initial gallbladder view. This frame was then given a PGS rating on the five-point, 1 (least inflamed) through 5 (most inflamed), scale, following the definitions outlined in Madni et al. [6]. Each subcomponent (adhesions, gallbladder appearance, distention, perforation, necrosis) of the PGS rating was also recorded. Timestamps for dissection tool created full-thickness gallbladder injuries during removal from the liver bed were also recorded.

Computer vision models

We trained two different computer vision models to classify the PGS of the representative image from each operative video. The first model, PGS-only, trained one neural network to classify PGS using the PGS labels alone. The second model, PGS-combo, trained two neural networks: one to classify level of adhesions and the other to classify gallbladder appearance. The PGS was then calculated from these subcomponents. Resnet50, a Convolutional Neural Network (CNN), was the visual network trained in all models, using the fastai deep learning library [11, 12]. Model training and performance was done with a tenfold cross-validation strategy. Stratified random sampling by PGS was performed to create the 10 folds. Each of the 10 folds was a 90:10 split of training to test images. Every network was then trained on each fold of the data. No data from the training sets was used in the test sets. For each training regimen, the network was trained for 35 epochs on training images that had undergone standard augmentation transforms with an initial learning rate of 2×10^{-3} followed by a discriminative learning rate. The initial learning rate was chosen with a learning rate finder [13].

Statistical analysis

We analyzed the effect of gallbladder inflammation, as represented by PGS, on three different outcomes using multilevel (mixed effects) regression models. We first looked at its effect on the magnitude (logarithm) of operative duration because factors that increase duration tend to do so exponentially. For example, a hole in the gallbladder leads to stone spillage, which then requires stone retrieval and irrigation. Likewise, an inflamed gallbladder is more difficult to grasp, which may make an already more difficult dissection harder to complete. This decision is further supported by literature that shows that surgical procedure times, across all varieties of procedures, follow a logarithmic distribution [7, 14]. We additionally looked at the effect of PGS on two other binary events: the likelihood that a surgeon would attain the CVS and the likelihood that a full-thickness injury (hole) would be created in the gallbladder during its removal.

Each model was similarly structured, with a varying (random) intercept for each surgeon, representing the outcome's value for that surgeon with a gallbladder PGS-1. Each surgeon was also assigned a varying (random) slope that represented the effect on the outcome for increases in PGS beyond 1. We used a multilevel varying intercepts and slopes approach grouped on surgeon to account for the effect of different surgeons' technique preferences on the outcomes, leveraging partial pooling to improve estimates for each surgeon, particularly those with few cases in the dataset. It also allowed us to model the correlation between a surgeon's outcome for a PGS-1 and the effect increases of PGS had on them. For example, if these were closely correlated, a surgeon that was fast at operating would be minimally affected by increases in PGS, while a slower surgeon would be more affected. We modeled PGS as an ordered categorical predictor, which allowed each incremental increase in PGS to have a different effect. As an example, going from a 1 to a 2 may have a smaller effect on the outcome than going from a 2 to a 3. Only cases from surgeons who contributed five or more cases was included. The full mathematical definition for each model is available in the *Supplement*.

All regression models were made and analyzed in R version 4.0.5 using the *rethinking* package version 2.13 [15, 16]. The *rethinking* package is an interface to Stan, a probabilistic programming language for specifying statistical models that uses Hamiltonian Monte Carlo sampling to provide Bayesian inference [17]. Every model used weakly regularizing priors. All Markov chains sampled well, with $\hat{R}$ values less than 1.01, and trace plots and rank histograms showed no evidence of biased posterior exploration nor divergent transitions [18]. Means and compatibility intervals (CI) were reported for regression coefficients. Compatibility intervals show the range for a parameter's value as seen across all values sampled in the model's Markov chains.

For each computer vision model, we calculated cross-validation performance, using Krippendorff's alpha for ordinal data as our agreement statistic. We chose Krippendorff's alpha since it assigns an increasing penalty to incorrect classifications as their distance from the truth increases. For example, mis-classifying a PGS-5 as a PGS-1 leads to a higher penalty than misclassifying it as a PGS-4. In general, an alpha of 0.8 or greater is considered highly reliable, and

an alpha between 0.667 and 0.800 is tentatively reliable [19]. The model's cross-validated performance was then compared to a second surgeon's PGS annotations. A confidence interval for the second surgeon's Krippendorff's alpha was using 10,000 bootstrap estimates. Calculation of Krippendorff's alpha was performed with the *irr* software package [20]. Graphics were generated with *ggplot2* and *ggdist* [21, 22]. All code for analyses is available online [23].

Results

Video information

Two hundred laparoscopic cholecystectomy videos were collected. In 196 videos, the entire case was recorded. 153 videos were performed by ten surgeons who had contributed five or more cases. The surgeons had a median [inter-quartile range (IQR)] of 10.6 (5.4–21.6) years of experience. Of the ten surgeons, four had minimally invasive fellowship training, four had trauma fellowship training, one was a hepatopancreaticobiliary surgeon, and one was a general surgeon. The dataset contained 42, 39, 42, 24, and 6 videos with a PGS of 1, 2, 3, 4, and 5, respectively. The median (IQR) laparoscopic duration was 38 (22–61) minutes. Surgeons obtained the CVS in 34% of cases and created inadvertent holes in the gallbladder during dissection from the liver bed 25% of the time.

Performance of computer vision models

We trained two computer vision models, PGS-only and PGS-combo, to classify PGS in a representative image from each case showing the initial gallbladder view. Using tenfold cross-validation, the two models agreed with the first surgeon's annotations, having a Krippendorff's alpha coefficient of 0.64 (95% CI 0.55–0.72) and 0.71 (95% CI 0.65–0.77), respectively. The PGS-combo model outperformed the PGS-only model due to improved classification of the underrepresented PGS-4 and PGS-5 gallbladders (Fig. 1). The PGS-combo model had comparable performance to that of the second surgeon annotator, whose Krippendorff's alpha coefficient was 0.82 (95% CI 0.75, 0.87) when compared to the first surgeon annotator.

Effect of gallbladder inflammation on operative outcomes

We analyzed the effect of gallbladder inflammation, as represented by PGS, on three different outcomes. We first analyzed its effect on laparoscopic case duration. On average across all surgeons and cases in our dataset, a PGS-2 resulted in minimal increases in operative duration, adding 2.7 (95% CI 0.3–7.0) minutes to an operation. Higher levels of inflammation caused large increases in operative duration, with a PGS-5 adding 16.9 (95% CI 4.4–33.9) minutes to the operation (Fig. 2). The PGS effect varied across surgeons, with the most affected surgeon experiencing an increase of 31.3 (95% CI 8.0–67.5) minutes when that surgeon operated

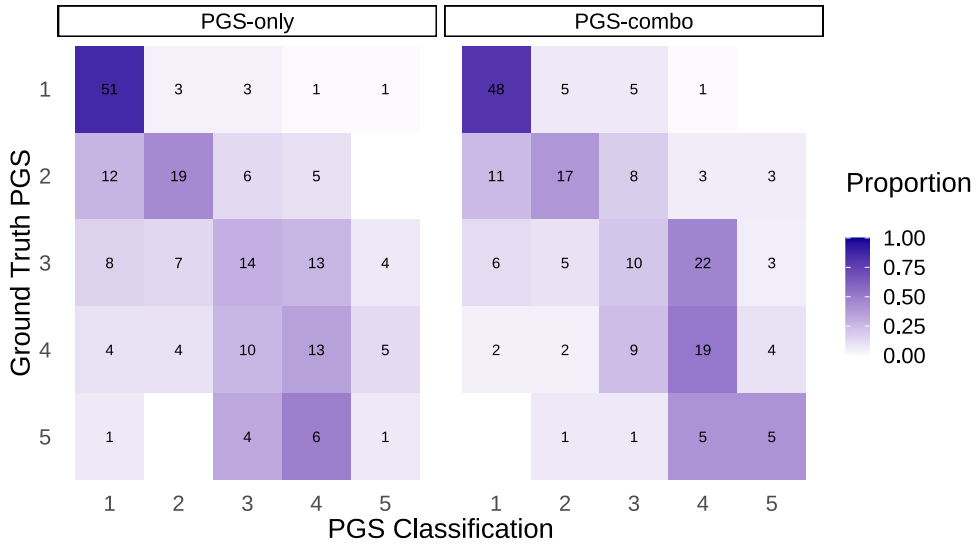


Fig. 1 Confusion matrix of the computer vision (CV) models, PGS-only and PGS-combo, classifying Parkland grading scale (PGS). PGS-combo was trained to classify PGS by recognizing its subcomponents (level of adhesions and gallbladder appearance), while PGS-only learned on the PGS alone. For each PGS, the proportion of images correctly classified is represented by the color in the heat map (white for low proportion; dark blue for high proportion). The PGS-combo model's improved classification of PGS-4 and PGS-5 led to performance gains over the PGS-only model (color figure online)

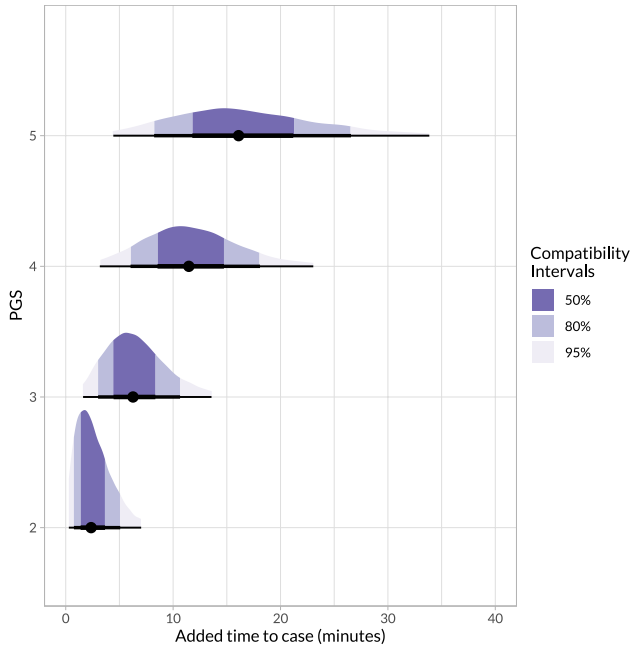


Fig. 2 Added inflammation increases the operative duration of laparoscopic cholecystectomy. Added time is compared to cases with minimal inflammation [Parkland grading scale (PGS) of 1]. The black dot and black bars represent the mean and compatibility interval for each value. Bar thickness, from most to least thick, and color, from dark purple to light purple, correspond to 50, 80, and 95% compatibility intervals. Compatibility intervals show the range for a parameter's value as seen across all values sampled in the model's Markov chains (color figure online)

on a maximally inflamed (PGS-5) gallbladder. There was minimal correlation ($\rho = 0.06$) between surgeons' cholecystectomy times for a minimally inflamed (PGS-1) gallbladder and the increase in operative duration they required with increasing gallbladder inflammation.

We next analyzed the effect of inflammation on the probability that a hole would be created in the gallbladder during its removal. Increases in inflammation resulted in minimal to weakly positive increase in the probability of a gallbladder hole from the baseline rate of 25% seen across all surgeons. PGS of 2, 3, 4, and 5 resulted in a percent increase in cases with an inadvertent hole (95% CI) of 2% (−2, 7), 5% (−4, 17), 8% (−6, 24) and 11% (−9, 34), respectively. As with the effect of inflammation on case duration, some surgeons were more affected by inflammation, with the most affected surgeon having a 28% (−2, 63) increase in the probability of a gallbladder hole for a PGS-5 (Fig. 3). There was minimal correlation ($\rho = -0.03$) between a surgeon's baseline probability of a gallbladder hole for a minimally inflamed gallbladder and the incremental effect of increasing inflammation.

Last, we analyzed the effect of inflammation on the probability that a surgeon would attain the CVS. Across all surgeons in the dataset, inflammation had no substantial effect on whether or not a surgeon attained the CVS, with a case where the PGS was 5 having an odds ratio (OR) of attaining the CVS, compared to a case with a PGS-1, of 0.57 (95% CI 0.09, 1.98). Unlike with operative duration and probability of a gallbladder hole, even on the most affected surgeon, inflammation had no substantial effect, with an OR for a PGS-5 compared to a PGS-1 of attaining the CVS of 0.58 (95% CI 0.05, 2.3) (Fig. 4). There was no correlation ($\rho = 0.00$) between surgeons' baseline probabilities of attaining the CVS for a minimally inflamed gallbladder versus those that were more inflamed.

Discussion

This study demonstrated the effect of varying levels of gallbladder inflammation on laparoscopic cholecystectomy operative duration and intra-operative events, including creation of inadvertent holes in the gallbladder and attainment of the CVS. Our statistical models built upon prior efforts and incorporated predictions personalized to each surgeon [4–6]. They additionally allowed for each increment in inflammation to have a different effect on the outcome. This flexible modeling demonstrated that going from a PGS-1 to PGS-2 gallbladder had little effect, in contrast to further increases in inflammation (Figs. 2, 3). Lastly, we demonstrated that gallbladder inflammation can reliably be identified by an AI computer vision model.

Operating room (OR) time is expensive and a limited resource. Large amounts of the literature have worked to optimize operative workflow to maximize OR utilization and minimize costs [24]. Orchestrating operative scheduling and workflow requires accurate predictions of case length. Traditionally, case duration has been predicted either from historical surgeon averages or statistical models that take into account the surgeon and procedure [7]. A more recent paper for laparoscopic cholecystectomy used pre-operative patient factors to improve prediction, however, predictive accuracy remained low [25]. Another paper used AI and computer vision to predict the next fifteen seconds of a case but did not try to predict overall case duration or the occurrence of any adverse intra-operative events [26]. Particularly for cases that vary substantially in length, such as cholecystectomy, these models fail to account for the intra-operative conditions that drive the variability in surgeons' times. We believe that every operation has a few "cornerstone" limiting factors upon which the majority of intra-operative variability rests. These cornerstones may range from the presence of intra-abdominal adhesions for gastrointestinal surgery or, in the case of cholecystectomy, the degree of gallbladder

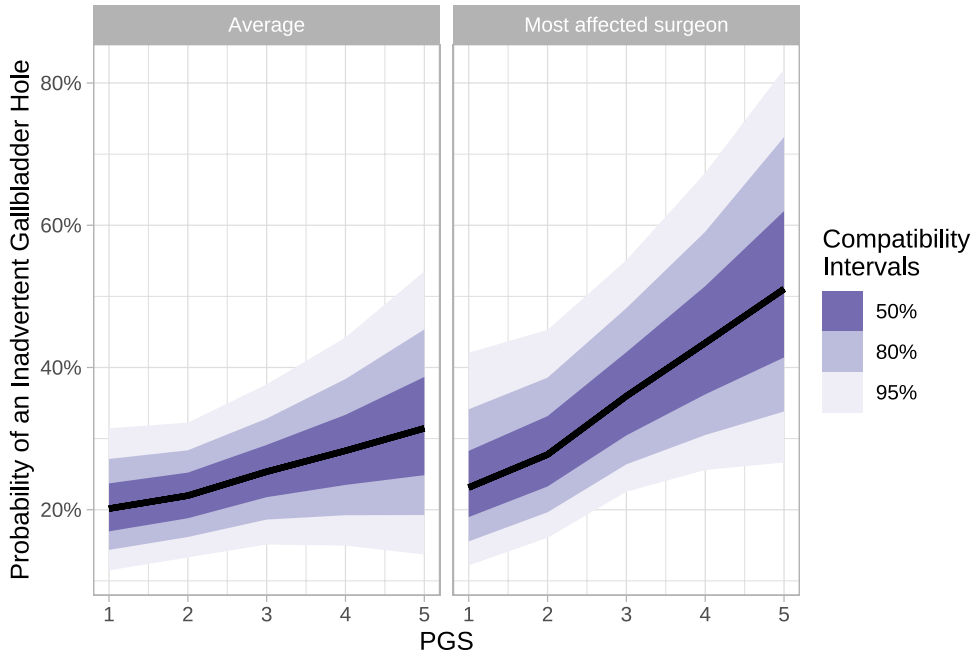


Fig. 3 For particular surgeons, added inflammation increases the probability of an inadvertent gallbladder hole. The black line and shade, from most to least dark, represent the mean and compatibility intervals (50, 80, and 95%) for each value. Compatibility intervals show the range for a parameter's value as seen across all values sampled in the model's Markov chains. Added inflammation, as measured by the Parkland grading scale (PGS), had little effect, on average across the dataset, towards increasing the probability of creating an inadvertent hole in the gallbladder during its removal from the liver bed. This effect, though, varied widely across surgeons, with some surgeons being particularly affected by increases in inflammation as seen in the right panel (color figure online)

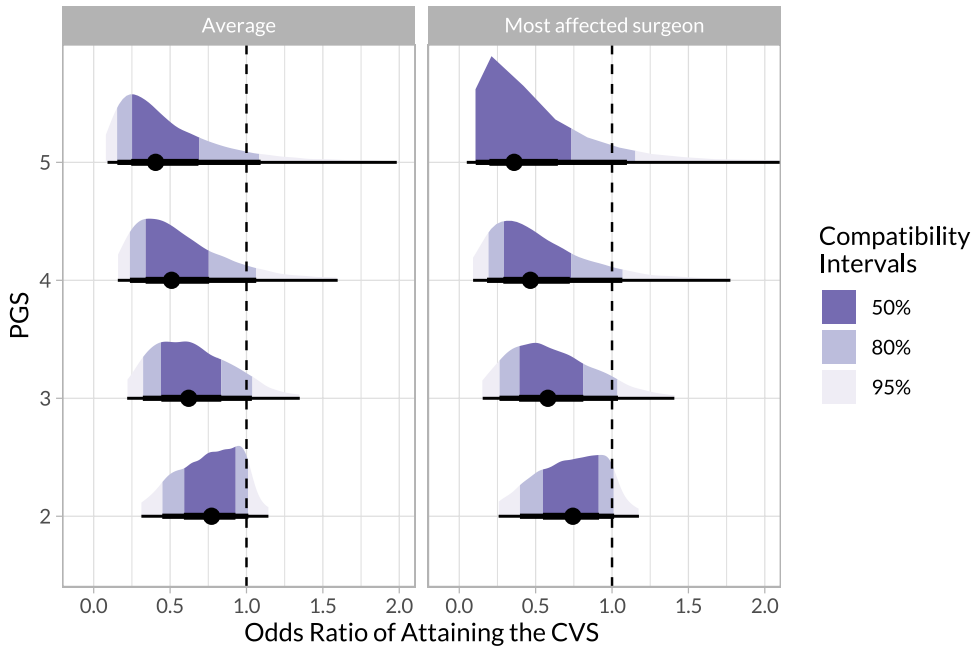


Fig. 4 Added inflammation had little effect on the odds that a surgeon attained the critical view of safety (CVS). Inflammation was measured using the Parkland grading scale (PGS) and compared to a PGS-1. The black dot and black bars represent the mean and compatibility interval for each value. Bar thickness, from most to least thick, and color, from dark purple to light purple, correspond to 50, 80, and 95 compatibility intervals. Compatibility intervals show the range for a parameter's value as seen across all values sampled in the model's Markov chains. Unlike the variation seen in the effect of PGS on case duration and gallbladder holes, even the most affected surgeon's likelihood of attaining the CVS was minimally impacted by increases in PGS (color figure online)

inflammation. Accurate operative time and event predictive models must account for these key factors. While a more recent operative time computer vision model, RSDNet, predicted remaining surgical duration at any point in the surgery, this model did not incorporate procedure-specific information, such as a cornerstone factor, and required the costly continuous running of an AI model during the case [27]. This contrasts with our model that takes a hybrid approach, where after processing a single image to determine the cornerstone limiting factor of cholecystectomy, it uses a more traditional multilevel regression model to rapidly output a predicted time.

Surgeons' attainment of the CVS was not changed by inflammation, which aligns with recent consensus recommendations that suggest the use of CVS for all laparoscopic cholecystectomies given no clear evidence to support adverse intra-operative events from attempts at attaining the CVS [28]. Severe inflammation of the hepatocystic triangle may preclude safe CVS attainment and therefore cause the surgeon to pursue alternative operative options (e.g., subtotal cholecystectomy, conversion to open). Since our study only included complete laparoscopic cholecystectomy videos, we were unable to investigate the effect of inflammation on creating hostile anatomy that forces a surgeon to pursue alternative operative options. Other studies though, have investigated this phenomenon and found a higher rate of conversion to open and partial cholecystectomy with more severe inflammation [6].

Beyond event prediction, the multilevel nature of the model also afforded a wealth of insight in per-surgeon metrics. For each event, it allowed for comparison of the surgeon to the average across all surgeons, both in how they perform when the gallbladder is minimally inflamed, but also in how they are affected by increments in inflammation. Surgeon performance on a gallbladder with minimal inflammation never correlated with the effect of increasing inflammation on any outcome. Therefore, surgeons' performances on a minimally inflamed gallbladders were not predictive of how they performed on maximally inflamed ones. For example, even though one surgeon had a similar rate of creating a gallbladder hole during a fifth of cases on a minimally inflamed gallbladder, for a maximally inflamed one, they, unlike the dataset average where inflammation minimally impacted the rate, created holes over half the time (Fig. 3). These personalized metrics would allow for focused feedback on particular aspects of the operation. Similar applications to resident assessment would be possible, which would allow for the transition from volume-based competency assessments to focused feedback that could better accelerate resident skill acquisition [29]. Recent advances in automated operative phase recognition and CVS assessment would allow for large-scale deployment of these metrics, using automated generation of data, which removes the substantial manual labor needed for video annotation [9, 30, 31].

This study is not without limitations. Inference on the effects of gallbladder inflammation represents the effects compatible with the videos and surgeons contained within the dataset and the statistical model's structure. Additionally, the inherent class imbalance noted with fewer examples of the most inflamed gallbladders, limits the precision of estimates for these cases. Despite all dataset cases being performed with a resident, no resident-specific data were included in the models. This decision was two-fold. First, attendings "select" the year of residents they work with through the inherent resident rotation schedules. They also exert control over the resident effect. For example, some attendings may allow for more resident independence in the case, and their times will reflect this. Second, the exact variable to include for a resident was unclear. Classifying residents by post-graduate year does not capture the inherent variation, both for a particular resident as the year progresses and inter-training year class variation. Our data collection also spanned multiple years, so a unique per-resident identifier would not help model precision as it would not account for a single resident's growth through residency. Additional considerations will be made in the future to determine how best to account for the impact of resident assistance on case metrics for attending surgeons.

In conclusion, an artificial intelligence computer vision model can reliably identify the degree of gallbladder inflammation. Using this information, we can predict cholecystectomy intra-operative course, including the laparoscopic duration, attainment of the CVS, and creation of inadvertent gallbladder holes. This automated assessment could immediately be useful for operating room workflow optimization and, in the near future, for targeted per-surgeon and per-resident feedback to accelerate acquisition of operative skills.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00464-022-09009-z.

Acknowledgements This work was supported by a 2018 research award from the Risk Management Foundation of the Harvard Medical Institutions Incorporated (CRICO/RMF), grant number 233456. The authors thank Caitlin E. Stafford, CCRP, for her assistance and support in research management, and Allison J. Navarrete-Welton, for her assistance in data collection.

Funding This study was funded by the Risk Management Foundation of the Harvard Medical Institutions Incorporated (CRICO/RMF), Grant Number 233456.

Declarations

Disclosures Drs. Ban, Hashimoto, Meireles, Rosman, and Ward receive research support from Olympus Corporation. Drs. Ban, Hashimoto,

Meireles, Rosman, and Ward have received research support from the Risk Management Foundation of the Harvard Medical Institutions Incorporated (CRICO/RMF). Dr. Hashimoto is a consultant for Johnson & Johnson, Activ Surgical, and Verily Life Sciences. Dr. Hashimoto has received research support from the Intuitive Foundation and the Society of American Gastrointestinal and Endoscopic Surgeons. Dr. Rosman receives research support from Toyota Research Institute (TRI). Dr. Meireles is a consultant for Medtronic and Olympus Corporation.

References

1. Sugrue M, Sahebally SM, Ansaloni L, Zielinski MD (2015) Grading operative findings at laparoscopic cholecystectomy—a new scoring system. World J Emerg Surg 10:14. https://doi.org/10.1186/s13017-015-0005-x
2. Nassar AHM, Ashkar KA, Mohamed AY, Hafiz AA (1995) Is laparoscopic cholecystectomy possible without video technology? Minim Invasive Ther 4:63–65. https://doi.org/10.3109/13645709509152757
3. Madni TD, Leshikar DE, Minshall CT, Nakonezny PA, Cornelius CC, Imran JB, Clark AT, Williams BH, Eastman AL, Minei JP, Phelan HA, Cripps MW (2018) The Parkland grading scale for Cholecystitis. Am J Surg 215:625–630. https://doi.org/10.1016/j.amjsurg.2017.05.017
4. Sugrue M, Coccolini F, Bucholc M, Johnston A, Manatakis D, Ioannidis O, Bonilauri S, Gachabayov M, Isik A, Ghnnam W, Shelat V, Aremu M, Mohan R, Montori G, Walędziak M, Pisarska M, Kong V, Strzałka M, Fugazzola P, Nita GE, Nardi M, Major P, Negoi I, Allegri A, Konstantoudakis G, Di Carlo I, Massalou D, D'Amico G, Solaini L, Ceresoli M, Bini R, Zielinski M, Tomasoni M, Litvin A, De Simone B, Lostoridis E, Hernandez F, Panyor G, Machain VGM, Pentara I, Baiocchi L, Ng KC, Ansaloni L, Sartelli M, Arellano ML, Savala N, Couse N, McBride S, Contributors from WSES (2019) Intra-operative gallbladder scoring predicts conversion of laparoscopic to open cholecystectomy: a WSES prospective collaborative study. World J Emerg Surg 14:12. https://doi.org/10.1186/s13017-019-0230-9
5. West Midlands Research Collaborative, Griffiths EA, Hodson J, Vohra RS, Marriott P, Katbeh T, Zino S, Nassar AHM (2019) Utilisation of an operative difficulty grading scale for laparoscopic cholecystectomy. Surg Endosc 33:110–121. https://doi.org/10.1007/s00464-018-6281-2
6. Madni TD, Nakonezny PA, Barrios E, Imran JB, Clark AT, Taveras L, Cunningham HB, Christie A, Eastman AL, Minshall CT, Luk S, Minei JP, Phelan HA, Cripps MW (2019) Prospective validation of the Parkland grading scale for Cholecystitis. Am J Surg 217:90–97. https://doi.org/10.1016/j.amjsurg.2018.08.005
7. Stepaniak PS, Heij C, Mannaerts GHH, de Quelerij M, de Vries G (2009) Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study. Anesth Analg 109:1232–1245. https://doi.org/10.1213/ANE.0b013e3181b5de07
8. Bellard F (2021) FFmpeg. https://ffmpeg.org/about.html. Accessed 21 Jun 2021
9. Ban Y, Rosman G, Ward T, Hashimoto D, Kondo T, Iwaki H, Meireles O, Rus D (2021) Aggregating long-term context for learning laparoscopic and robot-assisted surgical workflows. Accessed https://arxiv.org/abs/2009.00681
10. Strasberg SM, Hertl M, Soper NJ (1995) An analysis of the problem of biliary injury during laparoscopic cholecystectomy. J Am Coll Surg 180:101–125
11. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)
12. Howard J, Gugger S (2020) Fastai: a layered API for deep learning. Information 11:108. https://doi.org/10.3390/info11020108
13. Smith LN (2017) Cyclical learning rates for training neural networks. In: 2017 IEEE winter conference on applications of computer vision (WACV). pp 464–472
14. Strum DP, May JH, Vargas LG (2000) Modeling the uncertainty of surgical procedure times. Anesthesiology 92:1160–1167. https://doi.org/10.1097/00000542-200004000-00035
15. R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
16. McElreath R (2020) Rethinking: statistical rethinking book package. CRC Press, Boca Raton
17. Gabry J, Češnovar R (2020) cmdstanr: R interface to "CmdStan"
18. Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C (2021) Rank-normalization, folding, and localization: an improved R^ for assessing convergence of MCMC. Bayesian Anal 1:1–38. https://doi.org/10.1214/20-BA1221
19. Krippendorff K (2004) Content analysis: an introduction to its methodology, 2nd edn. Sage, Thousand Oaks
20. Gamer M, Lemon J, Singh IFP (2019) irr: various coefficients of interrater reliability and agreement. CRAN
21. Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York
22. Kay M (2021) ggdist: visualizations of distributions and uncertainty
23. Ward TM (2021) tmward/pgs: artificial intelligence prediction of cholecystectomy operative course from automated identification of gallbladder inflammation code. Accessed https://doi.org/10.5281/zenodo.5328655
24. Levine WC, Dunn PF (2015) Optimizing operating room scheduling. Anesthesiol Clin 33:697–711. https://doi.org/10.1016/j.anclin.2015.07.006
25. Thiels CA, Yu D, Abdelrahman AM, Habermann EB, Hallbeck S, Pasupathy KS, Bingener J (2017) The use of patient factors to improve the prediction of operative duration using laparoscopic cholecystectomy. Surg Endosc 31:333–340. https://doi.org/10.1007/s00464-016-4976-9
26. Ban Y, Rosman G, Ward T, Hashimoto D, Kondo T, Iwaki H, Meireles O, Rus D (2021) SUrgical PRediction GAN for events anticipation. Accessed https://arxiv.org/abs/2105.04642
27. Twinanda AP, Yengera G, Mutter D, Marescaux J, Padoy N (2019) RSDNet: learning to predict remaining surgery duration from laparoscopic videos without manual annotations. IEEE Trans Med Imaging 38:1069–1078. https://doi.org/10.1109/TMI.2018.2878055
28. The Prevention of Bile Duct Injury Consensus Work Group, Michael Brunt L, Deziel DJ, Telem DA, Strasberg SM, Aggarwal R, Asbun H, Bonjer J, McDonald M, Alseidi A, Ujiki M, Riall TS, Hammill C, Moulton C-A, Pucher PH, Parks RW, Ansari MT, Connor S, Dirks RC, Anderson B, Altieri MS, Tsamalaidze L, Stefanidis D (2020) Safe cholecystectomy multi-society practice guideline and state-of-the-art consensus conference on prevention of bile duct injury during cholecystectomy. Surg Endosc 34:2827–2855. https://doi.org/10.1007/s00464-020-07568-7
29. Ward TM, Mascagni P, Madani A, Padoy N, Perretta S, Hashimoto DA (2021) Surgical data science and artificial intelligence for surgical education. J Surg Oncol 124:221–230. https://doi.org/10.1002/jso.26496
30. Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Padoy N (2017) EndoNet: a deep architecture for recognition tasks on laparoscopic videos. IEEE Trans Med Imaging 36:86–97. https://doi.org/10.1109/TMI.2016.2593957

31. Mascagni P, Vardazaryan A, Alapatt D, Urade T, Emre T, Fiorillo C, Pessaux P, Mutter D, Marescaux J, Costamagna G, Dallemagne B, Padoy N (2021) Artificial Intelligence for Surgical Safety: Automatic Assessment of the Critical View of Safety in Laparoscopic Cholecystectomy Using Deep Learning. Ann Surg. https://doi.org/10.1097/SLA.0000000000004351

Supplemental Material

Formulas

Duration

$$\log(Duration_i) \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha_{sid[i]} + \beta_{sid[i]} * \sum_{j=0}^{PGS_i - 1} \delta_j$$

$$\begin{bmatrix} \alpha_{sid} \\ \beta_{sid} \end{bmatrix} \sim \text{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S}\right)$$

$$\alpha \sim \text{Normal}(0,1)$$

$$\beta \sim \text{Normal}(0,1.2)$$

$$\delta \sim \text{Dirichlet}(2)$$

$$\mathbf{S} = \begin{pmatrix} \sigma_a & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_a & 0 \\ 0 & \sigma_\beta \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$$\mathbf{R} \sim \text{LKJCorr}(4)$$

$$\sigma, \sigma_\alpha, \sigma_\beta \sim \text{Exponential}(1)$$

Gallbladder Holes

$$Hole_i \sim \text{Bernoulli}(p_i)$$

$$\text{logit}(p_i) = \alpha_{sid[i]} + \beta_{sid[i]} * \sum_{j=0}^{PGS_i - 1} \delta_j$$

$$\begin{bmatrix} \alpha_{sid} \\ \beta_{sid} \end{bmatrix} \sim \text{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S}\right)$$

$$\alpha \sim \text{Normal}(0,1.5)$$

$$\beta \sim \text{Normal}(0,0.75)$$

$$\delta \sim \text{Dirichlet}(2)$$

$$\mathbf{S} = \begin{pmatrix} \sigma_a & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_a & 0 \\ 0 & \sigma_\beta \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$$\mathbf{R} \sim \text{LKJCorr}(4)$$

$$\sigma_\alpha, \sigma_\beta \sim \text{Exponential}(1)$$

Critical View of Safety Attainment

$$CVS_i \sim \text{Bernoulli}(p_i)$$

$$\text{logit}(p_i) = \alpha_{\text{sid[i]}} + \beta_{\text{sid[i]}} * \sum_{j=0}^{PGS_i - 1} \delta_j$$

$$\begin{bmatrix} \alpha_{sid} \\ \beta_{sid} \end{bmatrix} \sim \text{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S}\right)$$

$$\alpha \sim \text{Normal}(0, 1.5)$$

$$\beta \sim \text{Normal}(0, 2)$$

$$\delta \sim \text{Dirichlet}(2)$$

$$\mathbf{S} = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$$\mathbf{R} \sim \text{LKJCorr}(4)$$

$$\sigma_\alpha, \sigma_\beta \sim \text{Exponential}(1)$$