

Original Article

Open Access



Artificial intelligence in laparoscopic cholecystectomy: does computer vision outperform human vision?

Runwen Liu^{1,2}, Jingjing An^{3,4}, Ziyao Wang^{1,2}, Jingye Guan⁵, Jie Liu⁵, Jingwen Jiang⁶, Zhimin Chen⁷, Hai Li⁷, Bing Peng^{1,2,8}, Xin Wang^{1,2}

¹Department of Pancreatic Surgery, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China.

²West China School of Medicine, Sichuan University, Chengdu 610041, Sichuan, China.

³Department of Operating Room, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China.

⁴West China School of Nursing, Sichuan University, Chengdu 610041, Sichuan, China.

⁵Chengdu Withai Innovation Technology Co., Ltd, Chengdu 610041, Sichuan, China.

⁶West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China.

⁷Department of Hepatobiliary Surgery, Chongzhou People's Hospital, Chengdu 610041, Sichuan, China.

⁸Department of 3rd General Surgery, Chengdu Shangjin Nanfu Hospital, Chengdu 610041, Sichuan, China.

#Authors contributed equally to this study.

Correspondence to: Dr. Xin Wang. Department of Pancreatic Surgery, West China Hospital, Sichuan University, No. 37, Guoxue Alley, Chengdu 610041, Sichuan, China. E-mail: hxwangxin2012@hotmail.com; Dr. Bing Peng, Department of Pancreatic Surgery, West China Hospital, Sichuan University, No. 37, Guoxue Alley, Chengdu 610041, Sichuan, China. E-mail: pengbing84@hotmail.com

How to cite this article: Liu R, An J, Wang Z, Guan J, Liu J, Jiang J, Chen Z, Li H, Peng B, Wang X. Artificial intelligence in laparoscopic cholecystectomy: does computer vision outperform human vision? *Art Int Surg* 2022;2:80-92. <https://dx.doi.org/10.20517/ais.2022.04>

Received: 25 Feb 2022 **First Decision:** 16 Mar 2022 **Revised:** 1 Apr 2022 **Accepted:** 19 Apr 2022 **Published:** 27 Apr 2022

Academic Editor: Andrew A. Gumbs **Copy Editor:** Tiantian Shi **Production Editor:** Tiantian Shi

Abstract

Background: The occurrence of biliary duct injury (BDI) after laparoscopic cholecystectomy (LC) remains 0.2-1.5%, which is largely caused by anatomic misidentifications. To solve this problem, we developed an artificial intelligence model, *SurgSmart*, and preliminarily verified its potential surgical guidance ability by comparing its performance with surgeons.

Methods: We prospectively collected 60 LC videos from November 2019 to August 2020 and enrolled 41 videos into the model establishment. Four important anatomic regions, namely cystic duct, cystic artery, common bile duct, and cystic plate, were annotated, and YOLOv3 (You Look Only Once), an object detection algorithm, was



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



applied to develop the model *SurgSmart*. To further evaluate its performance, comparisons were made among *SurgSmart*, trainees, and seniors (surgical experience in LC > 100).

Results: In total, 101,863 frames were extracted from videos, and 5533 video frames were selected, annotated, and used in model training. The mean average precision (mAP) of *SurgSmart* was 0.710. Comparative results show *SurgSmart* had significantly higher intersection-over-union (IoU) and accuracy (IoU \geq 0.5) in anatomy detection than those of seniors ($n = 36$) and trainees ($n = 32$) despite the existence of severe inflammation. Additionally, *SurgSmart* tended to correctly identify anatomic regions in earlier surgical phases than most of the seniors and trainees ($P < 0.001$).

Conclusions: *SurgSmart* is not only capable of accurately detecting and positioning anatomic regions in LC but also has better performance than that of the trainees and seniors in terms of individual still images and the whole set.

Keywords: Laparoscopic cholecystectomy, artificial intelligence, deep learning, computer vision, artificial intelligence-surgeon comparison

INTRODUCTION

Laparoscopic cholecystectomy (LC), one of the most frequent minimally invasive surgeries performed worldwide, has become a gold standard for patients with symptomatic gallstones^[1]. Although minimally invasive techniques have been developing for decades, biliary duct injury (BDI), which leads to re-interventions, increased mortality, and worsened quality of life^[2-4], remains one of the most severe complications for LC. To mitigate BDI, several safe intraoperative approaches, such as critical view of safety (CVS), intraoperative biliary imaging, and subtotal cholecystectomy, have been recommended in various circumstances during LC^[5,6]. Despite these preventive approaches, the overall incidence of BDI is still unchanged for many years and ranges from 0.2% to 1.5% globally^[7,8]. Such phenomenon is mainly due to the widely varying experience and anatomic localization ability of different surgeons. In an international survey involving over 600 surgeons, 72.3% of the surgeons had experienced BDI or near-misses, nearly half of whom acknowledged that the misrecognition of anatomic landmarks was the leading cause of BDI, especially between cystic duct and common bile duct^[9]. Hence, appropriately identifying pivotal structures including cystic duct, cystic artery, and common bile duct through anatomic guidance in real time is urgently needed.

Meanwhile, computer vision, a subfield of artificial intelligence (AI), has witnessed a dramatic resurgence in the past few years due to increases in computational power and the availability of big datasets^[10]. These technologies focus on image and video analysis, dealing with tasks such as object classification, detection, and segmentation^[11]. Meanwhile, medical images have greatly benefited from recent advances in computer vision. Many studies have indicated the promising results of computer vision in complex diagnostics such as dermatology^[12], radiology^[13], pathology^[14], and so on. These models were then mostly proved to be non-inferior or even superior to clinicians^[15,16]. Simultaneously, analysis of surgical videos, which contain a large amount of visible intraoperative information^[17], is another pivotal application scenario for computer vision. With the significant improvement in the efficiency and accuracy of the neural network^[18], more than 30 studies have demonstrated that computer vision could process a massive number of intraoperative videos to gain valuable information^[19]. For LC, computer vision could automatically identify the critical anatomic regions mentioned above^[20-22].

Therefore, in this study, we developed the computer vision model (*SurgSmart*) for LC and preliminarily verified its potential surgical guidance ability by comparing its performance with trainees and seniors.

METHODS

LC videos screening and the construction of anatomic dataset

From November 2019 to August 2020, we prospectively collected 60 LC videos in West China Hospital, Sichuan University, China. This study was approved by the Ethics Committee Board of our hospital. After excluding videos that were incomplete ($n = 1$), pixels or format incompatible ($n = 17$), or gallbladder gangrene ($n = 1$), 41 LC videos were included in the model development. Assigned by admission number and date of surgery, video frames were then extracted from videos included (one frame per second). Among these, ineligible frames (out-of-site, blurred, or duplicated) were excluded manually [Figure 1]. The remaining frames were then distributed and annotated by nine well-trained surgeons with over 100 LC video reviewing experiences^[23-25]. To ensure the accuracy and reliability of the annotation, the supervisor reviewed the annotators' work and made corrections if there was any mistake. Four important anatomic regions, namely cystic duct, cystic artery, common bile duct, and cystic plate, were annotated by bounding boxes [Supplementary Material 1]. All the work was conducted through Anaconda (Anaconda 2019, Inc, Austin, TX) in Python 3.6.5. NVIDIA (NVIDIA, Santa Clara, CA).

Algorithm development of SurgSmart

YOLOv3, an algorithm for object detection, was applied to develop *SurgSmart*. It consists of a residual network and a feature pyramid network. The residual network plays the role of improving processing efficiency, and the function of the feature pyramid network is feature extraction and multi-scale and multi-target detection [Supplementary Material 2]^[26]. Boosting scene understanding and multiple object detection ability of the model, the COCO dataset (a dataset provided by Microsoft used for image recognition) was used for pre-training and transfer learning^[27]. Meanwhile, multi-scale inputs for data augmentation were used to improve the performance of our model. Tesla V100 Graphic Processing Unit (GPU) was used for training, validation, and testing. Additionally, 80%, 10%, and 10% of the annotated frames from different videos were assigned into training, validation, and testing datasets, respectively [Figure 1]. The training and validation datasets came from the West China Hospital, and the test dataset came from its healthcare alliance.

SurgSmart model evaluation

Defined as the area of the intersection divided by the area of union [Supplementary Material 3a], intersection over union (IoU) was applied to determine the results of anatomic identification. The performance of *SurgSmart* was evaluated by precision {Precision = [(True Positive/(True Positive + False Positive))]} recall {Recall = [(True Positive/(True Positive + False Negative))]}, and average precision. Average precision (AP) is defined as the area under the precision-recall curve [Supplementary Material 3b]. Additionally, the mean average precision (mAP), which is the arithmetic mean of AP, was applied to evaluate the overall performance of *SurgSmart*.

Comparison between SurgSmart and trainees/seniors

To further evaluate the model's performance, we compared the anatomic recognitional ability of *SurgSmart*, trainees, and seniors. Trainees were defined as interns who had watched < 100 LC videos previously but without clinical experience, whereas seniors were defined as surgical fellows and attendings with > 100 LC operation experiences^[23,24]. The evaluation included marking three anatomic regions (cystic duct, cystic artery, and common bile duct) from 28 intraoperative still images from two LC videos, with corresponding time during video clips marked. Among them, 15 video pictures came from a video with low disease severity (Parkland Grade ≤ 2), and the others came from another video with a high inflammatory level (Parkland Grade ≥ 3).^[28] Anatomic identification results were evenly evaluated by dividing all the images into 7×5 equal parts and measured by IoU [Figure 2], which indicates correct identification if no less than 0.5. Apart from comparing the correct identification rate on the still image level, we also evaluated

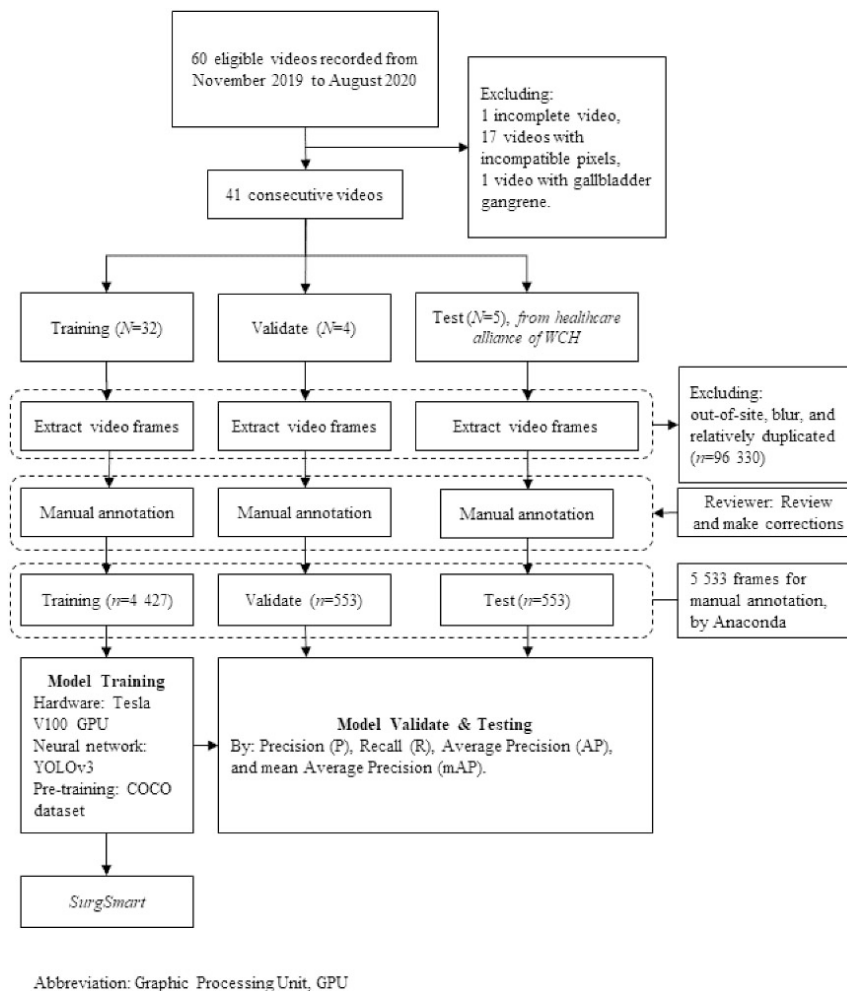


Figure 1. Inclusion and exclusion criteria of the video and the development of the model. GPU: graphic processing unit; WCH: West China Hospital.

the identification rate of three groups through the chronological level, which means identifying correctly one image in its whole set could be recognized as correctness. Furthermore, the time point (marked on each picture using the above method) of the earliest correct identification of each anatomic region was recorded and compared among the three groups.

Statistical analysis

We analyzed the comparison results through SPSS Statistics version 23.0 (IBM Corp., Armonk, NY). Cronbach's alpha and Kaiser-Meyer-Olkin and Bartlett's test of the results were used respectively to evaluate the reliability and construct validity of the test^[29,30]. Numeric data, IoU, were represented as ($\bar{x} \pm s_x$) and compared between *SurgSmart* and trainees/seniors through one-sample *t*-tests, and they were compared between trainees and seniors through independent samples *t*-test. For ranked data, we used Wilcoxon signed-rank test to compare detection time points and successful detection rate between *SurgSmart* and trainees/seniors and used the Mann-Whitney *U* test to compare them between trainees and seniors. The binominal test was used in comparisons of correct identification rates between *SurgSmart* and surgeons. P_a , P_b , and P' stand for *P*-value between *SurgSmart* and trainees, *SurgSmart* and seniors, trainees and seniors, respectively. $P < 0.05$ (both sides) was considered as statistically significant, and 95% of confidential intervals were all two-sided.

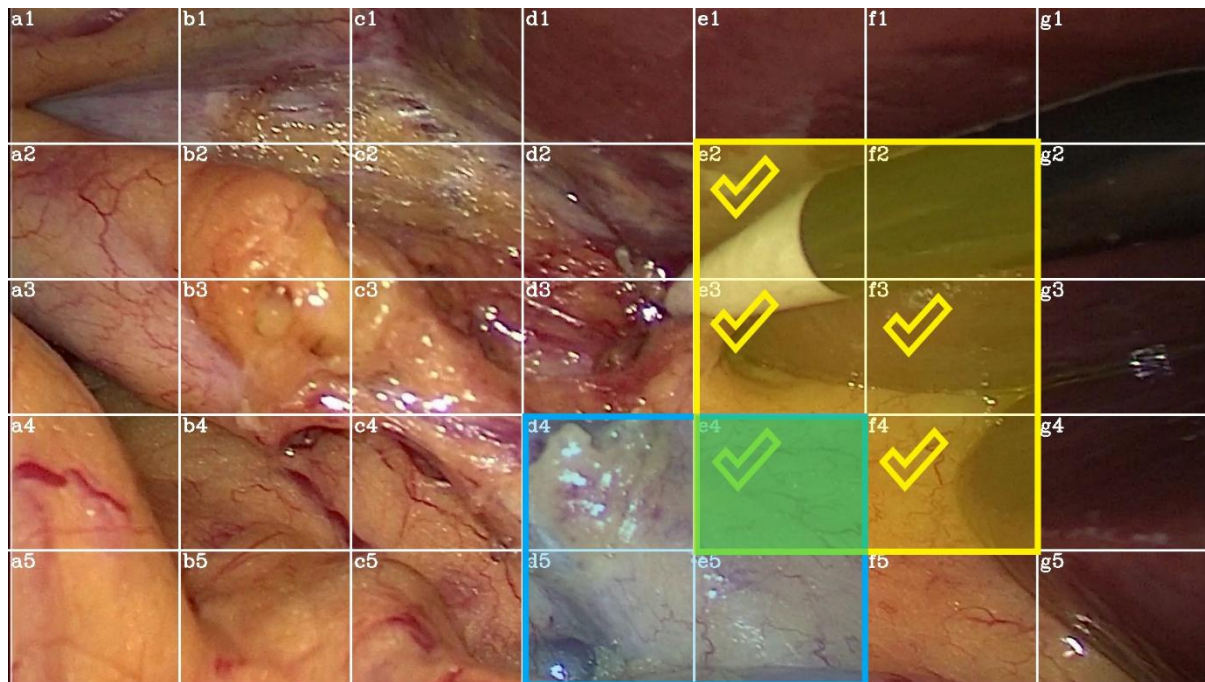


Figure 2. An example of judging detection IoU of human surgeon. The standard answer for common bile duct is “d4:e5” (shown in the blue rectangle with shadow). If the surgeon chose an answer(s), which could not form a rectangle, such as choosing e2, e3, f3, d4, and f4 (shown in yellow tick) simultaneously, it was converted into the horizontal rectangle to include the answer(s) below. Therefore, the IoU would be $\{[\text{Intersection (shown in green shadow)}/\text{Union}] = 1/9 \approx 0.111\}$ in this circumstance.

RESULTS

We extracted 101,863 frames from 41 LC videos, and 5533 frames were manually selected and annotated according to the criteria. There were 4427 (80%), 553 (10%), and 553 (10%) images randomly assigned to training, validation, and test datasets, respectively. The demographic characteristics of 41 patients who have undergone LC are shown in [Table 1](#), indicating most patients underwent elective surgery (37/41, 90.24%). Moreover, most patients were diagnosed with chronic calculous cholecystitis ($n = 34$, 82.93%) and only six (14.63%) patients with acute attacks.

For the performance of *SurgSmart*, the mAP of the four regions was 0.710. Specifically, the AP of the common bile duct, cystic duct, cystic artery, and the cystic plate was 0.855, 0.696, 0.504, and 0.787, respectively [[Table 2](#)].

After excluding invalid responses due to the nullification, 68 surgeons were involved in the AI-surgeon comparison. Among them, 32 participants were classified as trainees, and 36 surgeons were classified as seniors. The Cronbach’s alpha is 0.817 and 0.844 before and after excluding Questions 4, 9, 10, 18, and 27, respectively [[Supplementary Material 4](#)], showing good internal consistency. The validity of the test indicates mediocre construct validity, Kaiser-Meyer-Olkin Test and Bartlett’s Test results are 0.620 and 779.540 ($P < 0.001$), respectively, before the exclusion and 0.658 and 602.200 ($P < 0.001$), respectively, after such exclusion [[Supplementary Material 5](#)].

For each still image, the mean IoU of *SurgSmart*, trainees, and seniors were 0.561 , 0.299 ± 0.013 , and 0.450 ± 0.013 in cases without severe inflammation ($P_a, P_b, P' < 0.001$). The mean IoU of *SurgSmart*, trainees, and seniors were 0.646 , 0.275 ± 0.019 , and 0.407 ± 0.020 , respectively, in cases with severe inflammation ($P_a, P_b,$

Table 1. The demographic character of the video cases enrolled in model establishment

Factors	Statistical Results
Operation Age (Years) $\bar{x} \pm s$	51.75 \pm 11.09
Gender, N (%)	
Male	13 (31.71%)
Female	28 (68.29%)
Ethnicity, N (%) [#]	
Han	39 (95.12%)
Qiang	1 (2.44%)
Miao	1 (2.44%)
Surgery type, N (%)	
Elective	37 (90.24%)
Urgent	4 (9.76%)
Body Mass Index, kg/m ² , N (%)	
\leq 18.5	5 (12.20%)
18.5-23.9	23 (56.10%)
24.0-27.9	10 (24.39%)
\geq 28.0	3 (7.31%)
History of Hypertension, N (%)	
Yes	6 (14.63%)
No	35 (85.37%)
History of Hepatic Disease, N (%) [*]	
Yes	6 (14.63%)
No	35 (85.37%)
Hemoglobin (g/L), $\bar{x} \pm s$	136.78 \pm 14.45
White Blood Cells, 10 ⁹ /L, M(Q ₂₅ -Q ₇₅)	5.90 (4.50-7.75)
Alanine transaminase, IU/L, M(Q ₂₅ -Q ₇₅)	23 (16-38)
Triglyceride, mmol/L, M(Q ₂₅ -Q ₇₅)	1.30 (1.04-1.95)
Cholesterol, mmol/L, $\bar{x} \pm s$	4.77 \pm 0.89
hs-CRP (mg/L), M(Q ₂₅ -Q ₇₅)	1.43 (0.65-3.00)
APTT (s), $\bar{x} \pm s$	28.58 \pm 1.88
ASA Grading, N (%)	
1	6 (14.63%)
2	33 (80.49%)
3	2 (4.88%)
Pathology Results, N (%)	
Chronic Calculous Cholecystitis	34 (82.93%)
Acute Attack of Chronic Calculous Cholecystitis	6 (14.63%)
Chronic cholecystitis with cholesterol polyps	1 (2.44%)
Postoperative Hospitalization Days, M(Q ₂₅ -Q ₇₅)	1 (1-3)

APTT, activated partial thromboplastin time; hs-CRP, high-sensitive C-reactive protein; [#]All ethnicities listed are Asian; ^{*}Includes HBV hepatitis (n = 5) and fatty liver disease (n = 1).

Table 2. Testing results of SurgSmart

Metric	Mean Average Precision	Common bile duct	Cystic bed	Cystic duct	Cystic artery
Precision		0.861	0.842	0.747	0.637
Recall		0.905	0.848	0.784	0.632
Average Precision	0.710	0.855	0.787	0.696	0.504

$P' < 0.001$). Over half of the images witnessed statistically significant disparity between pairs of *SurgSmart*, trainees, and seniors ($P < 0.001$) [Supplementary Material 6]. Furthermore, according to the IoU curves shown in Figure 3, despite the IoU of three groups fluctuating at different levels among different images, the shape of curves had some similarities, especially in the case with severe inflammation.

Then, we also compared correct anatomic identification rates among the three groups. As shown in Table 3, in the case without inflammation, the accuracy of each structure recognition for *SurgSmart* on still images ranged from 60% to 100%, whereas those for trainees and seniors were 25.52%-27.34% and 40.74%-59.03%, respectively ($P_a, P_b, P' \leq 0.001$). At chronological images level, the correct identification rates of each anatomic region were 88.89%-96.37% for seniors, and those of trainees were 56.25%-68.75% ($P_a, P_b, P' < 0.001$). In cases with inflammation, the accuracy of anatomic regions for *SurgSmart* ranged from 66.67% to 100% on the still-image level, whereas those for trainees/seniors ranged from 8.33% to 75.93% ($P_a, P_b < 0.001$), the highest of which were those of common bile duct. On the chronological-images level, the correct identification rate of common bile duct for trainees and seniors was 68.75% and 94.44%, respectively ($P' = 0.006$). All these results demonstrate that the accuracy of *SurgSmart* in anatomic regional identification was significantly higher than that of seniors and trainees ($P_a, P_b < 0.001$).

Additionally, comparison results of the earliest surgical phases of correct identification are depicted in Figure 4. Even though seniors tended to correctly identify the anatomic regions in an earlier phase than trainees ($P' < 0.01$) except cystic artery in cases with severe inflammation, the performance of *SurgSmart* was superior to most of the surgeons regardless of experience ($P_a, P_b < 0.01$).

DISCUSSION

The ongoing development of computer vision technology, one of the areas of artificial intelligence, has made it possible to “recognize” images and videos with deep learning algorithms^[18]. Significant advances in these fields have resulted in AI achieving human-level object detection and scene recognition^[31,32]. As mentioned above, misrecognition of key anatomic regions such as the common bile duct is the main reason for BDI. Therefore, analyzing LC videos based on AI technologies may help us with clinical decision support and reduce iatrogenic injuries. However, AI-based surgical video analysis is still premature due to the complexity of the surgery and individual variance. In this study, we developed an AI-based anatomic recognition model, named *SurgSmart*, which could identify key anatomic regions of LC with a high degree of mAP (0.710). The anatomic recognitional performance comparison among *SurgSmart*, trainees, and seniors demonstrated that *SurgSmart* had a significantly higher IoU and correct anatomic identification rate than both trainees and seniors, regardless of disease severity. Moreover, *SurgSmart* also tended to correctly identify anatomic regions in earlier surgical phases during LC than most surgeons did.

Given that the misidentification of anatomic regions or landmarks during LC is the dominant factor contributing to BDI, correct recognition of these structures is fundamental for safe surgery^[5]. Although computer vision, a subtype of AI technology, has shown quite promising results for various pattern and object recognition tasks in medicine^[14,32], its application for real-time surgery is much more challenging since surgical videos have significantly more complexity in terms of background noise, anatomic regions, and various intraoperative information^[33]. In early research, Sato *et al.*^[34] explored the feasibility of ureter recognition of AI in 19 cases of laparoscopic hysterectomy. Although this study could not achieve satisfying results in detecting ureters using the Open-Source Computer Vision Library, the preliminary work still suggested the possibility of computer vision in assisting laparoscopic surgery. Later, Madad Zadeh *et al.*^[35] used another deep learning method called Mask R-CNN to develop an algorithm named SurgAI. Its detection accuracy and segmentation accuracy of ureter were 97% and 84.5%, respectively, which further

Table 3. Comparison of successful detection rate of anatomic landmarks (threshold of IoU: 0.5)

With severe inflammation	Category of successful detection Rate	Anatomic Parts	SurgSmart	Trianees	P_a	Seniors	P_b	P'	
No	Still images	Cystic duct	4/4 (100.00%)	35/128 (27.34%)	--	85/144 (59.03%)	--	< 0.001	
		Cystic artery	5/6 (83.33%)	49/192 (25.52%)	< 0.001^a	88/216 (40.74%)	< 0.001^a	0.001	
		Common bile duct	3/5 (60.00%)	52/160 (32.50%)	< 0.001^a	96/180 (53.33%)	< 0.001^a	< 0.001	
	Chronological	Cystic duct	--	18/32 (56.25%)	--	35/36 (96.37%)	--	< 0.001	
		Cystic artery	--	22/32 (68.75%)	--	35/36 (96.37%)	--	0.002	
		Common bile duct	--	21/32 (65.62%)	--	32/36 (88.89%)	--	0.002	
	Yes	Still images	Cystic duct	2/3 (66.67%)	10/96 (10.42%)	< 0.001^a	21/108 (19.44%)	< 0.001^a	0.074
			Cystic artery	2/2 (100.00%)	8/64 (12.50%)	--	6/72 (8.33%)	--	0.427
			Common bile duct	3/3 (100.00%)	47/96 (48.96%)	--	82/108 (75.93%)	--	< 0.001
Choronoglogical		Cystic duct	--	9/32 (28.12%)	--	21/36 (58.33%)	--	0.013	
		Cystic artery	--	8/32 (25.00%)	--	6/36 (16.67%)	--	0.4	
		Common bile duct	--	22/32 (68.75%)	--	34/36 (94.44%)	--	0.006	

indicated the feasibility of AI-assisted anatomy recognition using the AI technique.

To mitigate the BDI rate of LC by AI technologies, several institutes have carried out relevant preliminary studies, some of which demonstrated promising results: Tokuyasu *et al.*^[36] developed and validated an AI-assisted anatomy recognition system for LC. In this study, YOLOv3, an algorithm of objects detection, was applied to identify anatomic regions (such as cystic duct and common bile duct) and landmarks recognition [such as the lower edge of liver segment 4 (S4) and Rouviere's sulcus]. Although the AP for each structure or landmark showed poor results, ranging from 0.074 to 0.320, the valid anatomic identification was confirmed on 22 of the 23 videos, demonstrating its great potential to mitigate BDI. In our study, four anatomic structures, directly or indirectly associated with CVS, were selected for model development. After amplifying the training datasets and optimizing the YOLOv3 model by sample equalization, the mAP of key anatomic regions was significantly increased to 0.710, even in cases with severe inflammation^[28]. Meanwhile, other studies have explored anatomic evaluation approaches, such as safe/dangerous zone identification. For instance, Madani *et al.*^[33] developed and validated the performance of an AI algorithm that could identify safe/dangerous zones based on the location within the hepatocystic triangle, with the potential for real-time guidance during LC. In addition, automatic evaluation of CVS, which is a gold standard of anatomic processing during LC, was also applied to reduce BDI^[5]. Korndorffer *et al.*^[21] developed an auto-evaluating model for CVS, which showed promising results, especially in cases with severe inflammation. Moreover, according to the latest research, Mascagni *et al.*^[37] developed a deep learning model to automatically assess CVS during LC with a mean average precision of 71.9%, suggesting AI could be able to assess CVS criteria. Therefore, with the help of continuous dataset training and ongoing breakthroughs in computer algorithms, combined with multiple methods, such as anatomy detection/segmentation, CVS

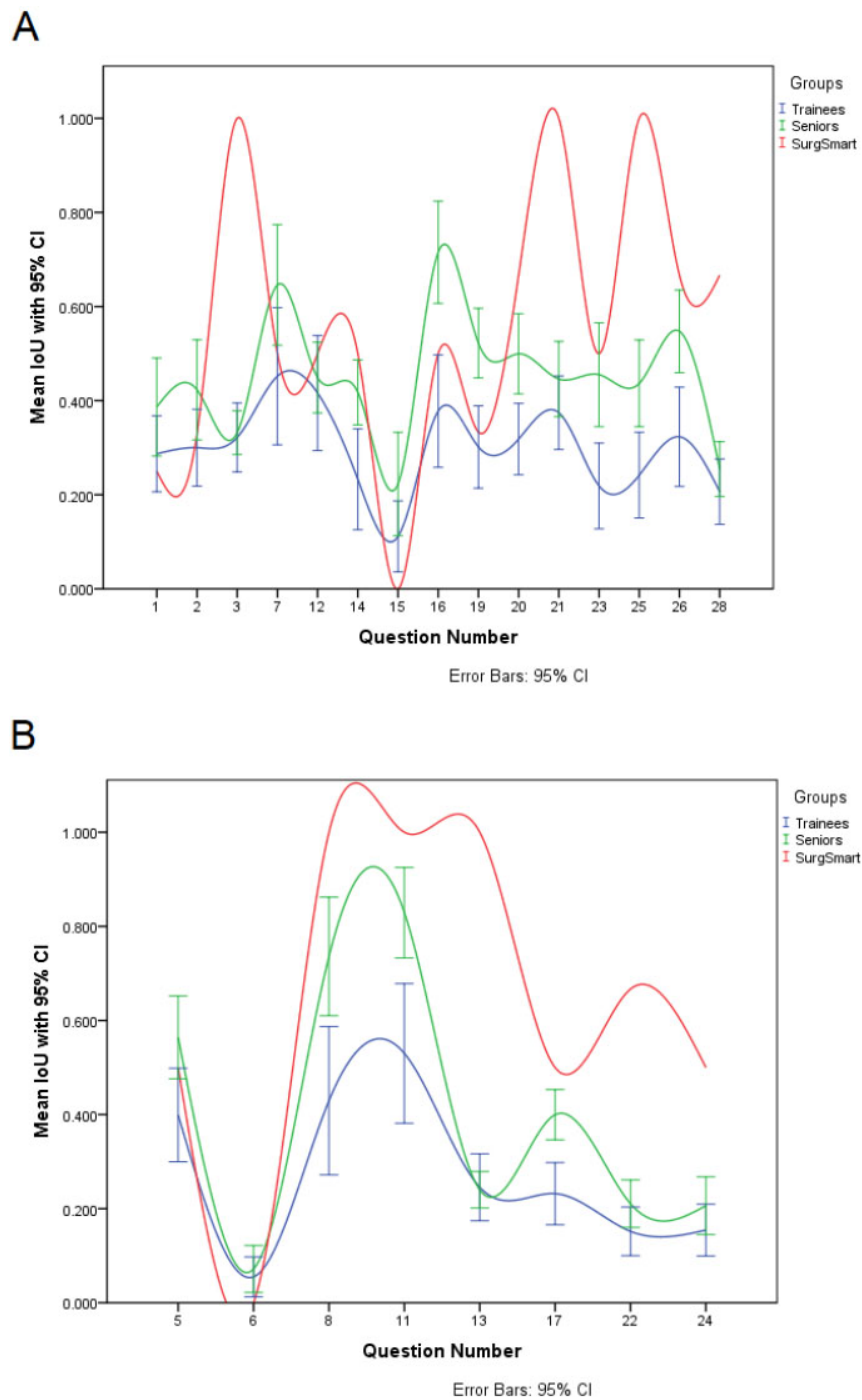


Figure 3. Comparison of overall detection results in frames of LC with and without severe inflammation: (A) frames from a low-severity case (Parkland Grade ≤ 2); and (B) frames from a high-severity case (Parkland Grade ≥ 3). CI, Coefficient Interval; IoU, Intersection over Union. The question numbers were arranged arbitrarily. Questions 4, 9, 10, 18, and 27 were excluded.

evaluation, safe-zones guidance, and more, AI systems will be applied for supporting the clinical decision and ensuring the safety during surgery.

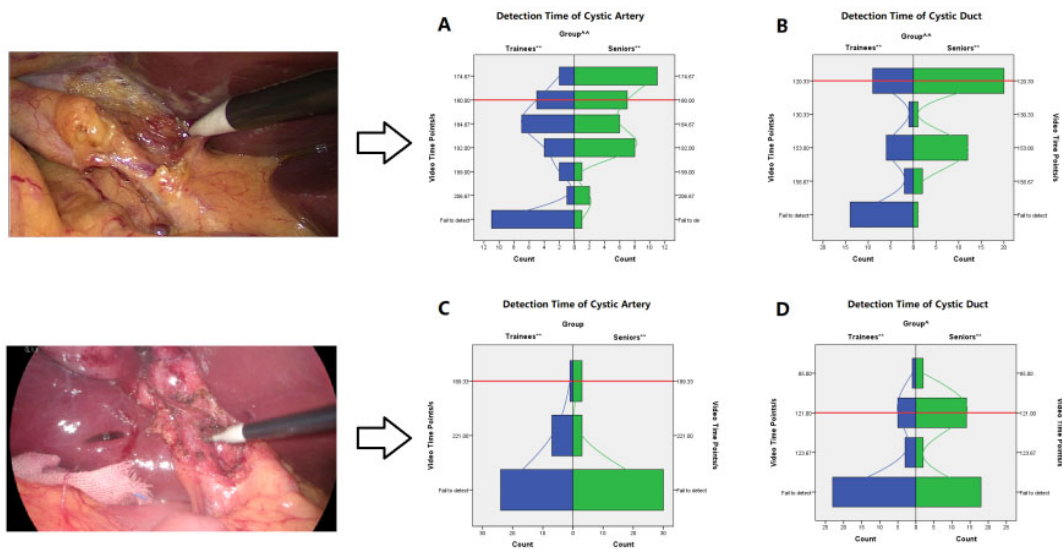


Figure 4. Comparison of detection time point of human surgeons and *SurgSmart* towards cystic duct and artery with and without severe inflammation: (A, B) for a low-severity case (Parkland Grade ≤ 2); and (C, D) for a high-severity case (Parkland Grade ≥ 3). * $P < 0.05$ between trainees/seniors and *SurgSmart*; ** $P < 0.01$ between trainees/seniors and *SurgSmart*; ^ $P < 0.05$ between trainees and seniors; ^^ $P < 0.01$ between trainees and seniors. The red reference line stands for the detection time point of *SurgSmart*.

To be noted, the significantly varied performance of AI recognition among different anatomic regions probably lies in their feature. To begin with, the size of the anatomic region has a significant influence on AI performance. In the studies conducted by Madad Zadeh *et al.*^[35], the accuracy of large anatomic regions such as the uterus was 84.5% in contrast with those with small anatomy regions such as ovaries, only 29.6% in the SurgAI algorithm, indicating larger objects always have better recognition effects. Furthermore, a clear and regular boundary of anatomic regions will be conducive to deep learning. Structures such as the gallbladder and liver have significantly higher IoU when compared to dissected hepatocystic triangle and cystic plate^[33,37]. Additionally, some anatomic regions with various intraoperative exposure conditions are difficult for AI recognition. In our study, the exposure state of “dynamic” structures such as cystic duct and cystic artery change constantly during the hepatocystic mobilization, while those of “non-dynamic” regions such as liver and common bile duct almost remain unchanged, leading to the variance of recognition performance in terms of AP. Thus, balanced training datasets and individualized algorithms for different anatomic regions might be essential for optimizing the overall capability of AI systems.

In 2017, Esteva *et al.*^[12] suggested that deep learning achieved performance on par with all board-certificated dermatologists in terms of skin cancer diagnosis, indicating that AI can classify skin cancer comparable to experts. Later, several studies demonstrated that the AI systems performed at physician-like levels in lesions identification and classification, such as breast cancer detection in mammography^[15], lung cancer diagnosis in CT^[16], and others. Nevertheless, it is estimated that a 1 min high-quality surgical clip contains 25 times as many data as a high-resolution CT image^[38], bringing great difficulty to the application of AI in surgery. To further verify the AI’s anatomic identifying capability, we compared the performance among *SurgSmart*, trainees, and seniors. Unexpectedly, the model’s overall IoU and correct identification rate are significantly higher than those of seniors and trainees on still images level. Although severe inflammation will seriously affect the anatomic identification for surgeons regardless of experience, it has no impact on the performance of *SurgSmart*. Furthermore, *SurgSmart* can correctly identify anatomic sites at earlier surgical phases than most surgeons. These results suggest that AI does well in identifying objects or scenes based on some characteristics hidden in still images, which are difficult to be detected or noticed by humans. Conversely,

humans need more relational and logical information obtained through visual inspection, sequenced information, exploring by instruments, and the sense of touch, to identify complex objects. Hence, the omission of the latter three approaches during the experiment probably led to the low effective anatomic identification rate. Noticeably, the effective identification rate of the key region, such as the common bile duct in the senior group, approached over 90% when considering sequenced information. Therefore, the dependency of AI and humans on various information is quite different, leading to the potential necessity of intelligent guidance, especially for novice surgeons.

Interestingly, the shape of IoU curves in [Figure 3](#) is very similar among the three groups, indicating the anatomic region that is complex for surgeons, as in the cystic artery, also impacts the recognition results of AI. Despite the difference between AI and trainees/seniors in terms of the depth of understanding of still images, some similarities might exist in their learning process. Besides, the performance of *SurgSmart* in images with severe inflammation tends to be better than those without severe inflammation, which is similar to the result of a recent CVS evaluation study performed by Korndorffer *et al.*^[21] Therefore, utilizing the learning mechanism in humans may further improve the capability of AI in the future.

There are several limitations to the study. First, all surgical videos and images used for algorithm training came from limited institutes with restricted external access due to the hospital regulations. The current system, *SurgSmart*, may not be fully generalizable to other hospitals. To solve this problem, we have launched a multi-center project named “LC10000”, which will collect over 10,000 videos of LC in the next few years prospectively to expand the diversity of the dataset. Second, the comparison between the AI and surgeons only consisted of two cases, which consisted of 28 questions of still images required to select key areas of anatomic regions and might not directly translate to superior detection of anatomy during “live” surgery. Nonetheless, some participants felt overwhelmed and lassitude, which may also occur in the later stage of operations while facing a relentless set of tests, leading to poor results, especially near the end of the test. Thus, the distortion of results will generate if we expand the number of cases to incessantly evaluate the anatomic detection ability of surgeons, who would relentlessly answer many questions. Additionally, due to the low fault-tolerance rate of the surgical process, which means small mistakes may bring disastrous consequences to patients, it is difficult to achieve this by automated anatomy recognition only. Hence, combining various approaches, such as CVS evaluation, dangerous zone indication, and anatomic structure identification, would be helpful to enrich the AI system in the field of surgery.

In conclusion, *SurgSmart* can detect and localize key anatomic regions of LC accurately. In a series of still images, its anatomic locating ability is significantly more precise than those of surgeons. Although there is still a long way to go before *SurgSmart* can be utilized clinically, with the expansion of datasets and developments in the algorithm, we believe that a well-trained AI system will be born soon, translating the advantages of AI into benefits for patients and surgeons.

DECLARATIONS

Acknowledgements

We deeply grateful to the contribution of all the authors, and to the technical supports of Yuxian Wang, Zunyu Liu, and Shaonan Wu.

Authors' Contributions

Study conception and design: Liu R, An J, Wang Z, Jiang J, Chen Z, Li H, Peng B, Wang X

Acquisition of data: An J, Guan J, Liu J, Wang Z, Chen Z, Li H, Wang X

Analysis and interpretation of data: Liu R, Guan J, Liu J, Jiang J, Chen Z, Li H

Drafting of manuscript: Liu R, Wang Z, Guan J, Liu J, Wang X

Critical revision: An J, Jiang J, Peng B, Wang X

Availability of data and materials

Due to the administrative issues of West China Hospital and its allies, the relevant data is temporarily not available to the public.

Financial support and sponsorship

This work is supported by key project of Health Commission of Sichuan Province, China (20ZD003).

Conflicts of interest

All authors declared that there is no conflict of interest.

Ethical approval and consent to participate

This study was approved by the Ethics Committee Board of West China HoWest China Hospital, Sichuan University, China. All procedures performed in studies involving human participants had been performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2022.

REFERENCES

1. Soper NJ, Stockmann PT, Dunnegan DL, Ashley SW. Laparoscopic cholecystectomy. The new “gold standard”? *Arch Surg* 1992;127:917-21; discussion 921. [DOI PubMed](#)
2. Schwaitzberg SD, Scott DJ, Jones DB, et al. Threefold increased bile duct injury rate is associated with less surgeon experience in an insurance claims database: more rigorous training in biliary surgery may be needed. *Surg Endosc* 2014;28:3068-73. [DOI PubMed](#)
3. Törnqvist B, Strömberg C, Akre O, Enochsson L, Nilsson M. Selective intraoperative cholangiography and risk of bile duct injury during cholecystectomy. *Br J Surg* 2015;102:952-8. [DOI PubMed](#)
4. Barrett M, Asbun HJ, Chien HL, Brunt LM, Telem DA. Bile duct injury and morbidity following cholecystectomy: a need for improvement. *Surg Endosc* 2018;32:1683-8. [DOI PubMed](#)
5. Brunt LM, Deziel DJ, Telem DA, et al; the Prevention of Bile Duct Injury Consensus Work Group. Safe cholecystectomy multi-society practice guideline and state of the art consensus conference on prevention of bile duct injury during cholecystectomy. *Ann Surg* 2020;272:3-23. [DOI PubMed](#)
6. Lilley EJ, Scott JW, Jiang W, et al. Intraoperative cholangiography during cholecystectomy among hospitalized medicare beneficiaries with non-neoplastic biliary disease. *Am J Surg* 2017;214:682-6. [DOI PubMed](#)
7. Törnqvist B, Strömberg C, Persson G, Nilsson M. Effect of intended intraoperative cholangiography and early detection of bile duct injury on survival after cholecystectomy: population based cohort study. *BMJ* 2012;345:e6457. [DOI PubMed PMC](#)
8. Fong ZV, Pitt HA, Strasberg SM, et al; California Cholecystectomy Group. Diminished survival in patients with bile leak and ductal injury: management strategy and outcomes. *J Am Coll Surg* 2018;226:568-576.e1. [DOI PubMed PMC](#)
9. Iwashita Y, Hibi T, Ohyama T, et al. Delphi consensus on bile duct injuries during laparoscopic cholecystectomy: an evolutionary cul-de-sac or the birth pangs of a new technical framework? *J Hepatobiliary Pancreat Sci* 2017;24:591-602. [DOI PubMed](#)
10. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med* 2019;25:24-9. [DOI PubMed](#)
11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44. [DOI PubMed](#)
12. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-8. [DOI PubMed PMC](#)
13. Cicero M, Bilbily A, Colak E, et al. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. *Invest Radiol* 2017;52:281-7. [DOI PubMed](#)
14. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in

- retinal fundus photographs. *JAMA* 2016;316:2402-10. DOI PubMed
15. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. *J Natl Cancer Inst* 2019;111:916-22. DOI PubMed PMC
 16. Xie Y, Zhang J, Xia Y. Semi-supervised adversarial model for benign-malignant lung nodule classification on chest CT. *Med Image Anal* 2019;57:237-48. DOI PubMed
 17. Grenda TR, Pradarelli JC, Dimick JB. Using surgical video to improve technique and skill. *Ann Surg* 2016;264:32-3. DOI PubMed PMC
 18. Tan Mingxing, Le Quoc V. EfficientNet: rethinking model scaling for convolutional neural networks. 2019. Available from: <http://proceedings.mlr.press/v97/tan19a/tan19a.pdf> [Last accessed on 22 Apr 2022]. DOI
 19. Anteby R, Horesh N, Soffer S, et al. Deep learning visual analysis in laparoscopic surgery: a systematic review and diagnostic test accuracy meta-analysis. *Surg Endosc* 2021;35:1521-33. DOI PubMed
 20. Yamazaki Y, Kanaji S, Matsuda T, et al. Automated surgical instrument detection from laparoscopic gastrectomy video images using an open source convolutional neural network platform. *J Am Coll Surg* 2020;230:725-732.e1. DOI PubMed
 21. Korndorffer JR Jr, Hawn MT, Spain DA, et al. Situating artificial intelligence in surgery: a focus on disease severity. *Ann Surg* 2020;272:523-8. DOI PubMed
 22. Twinanda AP, Yengera G, Mutter D, Marescaux J, Padoy N. RSDNet: Learning to predict remaining surgery duration from laparoscopic videos without manual annotations. *IEEE Trans Med Imaging* 2019;38:1069-78. DOI PubMed
 23. Abdelrahman T, Long J, Egan R, Lewis WG. Operative experience vs. competence: a curriculum concordance and learning curve analysis. *J Surg Educ* 2016;73:694-8. DOI PubMed
 24. Marchi D, Esposito MG, Gentile IG, Gilio F. Laparoscopic cholecystectomy: training, learning curve, and definition of expert. in: Agresta F, Campanile FC, Vettoretto N, editors. *Laparoscopic cholecystectomy*. Cham: Springer International Publishing; 2014. pp. 141-7.
 25. Maier-Hein L, Wagner M, Ross T, et al. Heidelberg colorectal data set for surgical data science in the sensor operating room. *Sci Data* 2021;8:101. DOI PubMed PMC
 26. Liu C, Guo Y, Li S, Chang F. ACF based region proposal extraction for YOLOv3 network towards high-performance cyclist detection in high resolution images. *Sensors (Basel)* 2019;19:E2671. DOI PubMed PMC
 27. Lin Tsung-Yi, Maire Michael, Belongie Serge, et al. , editors. *Microsoft COCO: common objects in context 2014*; Cham: Springer International Publishing.
 28. Madni TD, Leshikar DE, Minshall CT, et al. The Parkland grading scale for cholecystitis. *Am J Surg* 2018;215:625-30. DOI PubMed
 29. Movafegh F, Abbaszadeh A, Rassouli M, Lotfi MS, Nasiri M, Makhlesi S. Development and validation of the Iranian version of the patient privacy and confidentiality scale. *Indian J Med Ethics* 2021;VI:1-13. DOI PubMed
 30. Kristen J. , Yoon-Soo P., Emil P. Are your assessment scores and feedback reliable? A statistical review for the surgical educator 2021. Available from: <https://www.facs.org/education/division-of-education/publications/rise/articles/assessment-scores> [Last accessed on 22 Apr 2022]. DOI
 31. Barisoni L, Lafata KJ, Hewitt SM, Madabhushi A, Balis UGJ. Digital pathology and computational image analysis in nephropathology. *Nat Rev Nephrol* 2020;16:669-85. DOI PubMed PMC
 32. Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence-assisted gastroenterology- promises and pitfalls. *Am J Gastroenterol* 2019;114:422-8. DOI PubMed
 33. Madani A, Namazi B, Altieri MS, et al. Artificial intelligence for intraoperative guidance: using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy. *Ann Surg* 2020. DOI PubMed PMC
 34. Sato M, Koizumi M, Nakabayashi M, et al. Computer vision for total laparoscopic hysterectomy. *Asian J Endosc Surg* 2019;12:294-300. DOI PubMed
 35. Madad Zadeh S, Francois T, Calvet L, et al. SurgAI: deep learning for computerized laparoscopic image understanding in gynaecology. *Surg Endosc* 2020;34:5377-83. DOI PubMed
 36. Tokuyasu T, Iwashita Y, Matsunobu Y, et al. Development of an artificial intelligence system using deep learning to indicate anatomical landmarks during laparoscopic cholecystectomy. *Surg Endosc* 2021;35:1651-8. DOI PubMed PMC
 37. Mascagni P, Vardazaryan A, Alapatt D, et al. Artificial intelligence for surgical safety: automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning. *Ann Surg* 2020. DOI PubMed
 38. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg* 2018;268:70-6. DOI PubMed PMC