Situating Artificial Intelligence in Surgery

A Focus on Disease Severity

James R. Korndorffer Jr, MD, Mary T. Hawn, MD, David A. Spain, MD, Lisa M. Knowlton, MD, Dan E. Azagury, MD, Aussama K. Nassar, MD, James N. Lau, MD, Katherine D. Arnow, MS, Amber W. Trickey, PhD, MS, and Carla M. Pugh, MD, PhD✉

Objectives: Artificial intelligence (AI) has numerous applications in surgical quality assurance. We assessed AI accuracy in evaluating the critical view of safety (CVS) and intraoperative events during laparoscopic cholecystectomy. We hypothesized that AI accuracy and intraoperative events are associated with disease severity.

Methods: One thousand fifty-one laparoscopic cholecystectomy videos were annotated by AI for disease severity (Parkland Scale), CVS achievement (Strasberg Criteria), and intraoperative events. Surgeons performed focused video review on procedures with $\geq$1 intraoperative events (n = 335). AI versus surgeon annotation of CVS components and intraoperative events were compared. For all cases (n = 1051), intraoperative-event association with CVS achievement and severity was examined using ordinal logistic regression.

Results: Using AI annotation, surgeons reviewed 50 videos/hr. CVS was achieved in $\leq$10% of cases. Hepatocystic triangle and cystic plate visualization was achieved more often in low-severity cases ($P < 0.03$). AI-surgeon agreement for all CVS components exceeded 75%, with higher agreement in high-severity cases ($P < 0.03$). Surgeons agreed with 99% of AI-annotated intraoperative events. AI-annotated intraoperative events were associated with both disease severity and number of CVS components not achieved. Intraoperative events occurred more frequently in high-severity versus low-severity cases (0.98 vs 0.40 events/case, $P < 0.001$).

Conclusions: AI annotation allows for efficient video review and is a promising quality assurance tool. Disease severity may limit its use and surgeon oversight is still required, especially in complex cases. Continued refinement may improve AI applicability and allow for automated assessment.

Keywords: artificial intelligence, cholecystectomy, critical view of safety, quality assurance, video-based assessment

(*Ann Surg* 2020;272:523–528)

From the Department of Surgery, Stanford University, Stanford, CA.
✉cpugh@stanford.edu.

James R. Korndorffer, Jr.: substantial contributions to conception, design, acquisition, analysis, and interpretation of data; as well as drafting the article and revising it critically for important intellectual content. Gave final approval.
Mary T. Hawn: substantial contributions to conception and design, analysis, and interpretation of data; as well as drafting the article and revising it critically for important intellectual content. Gave final approval.
David A. Spain: substantial contributions to conception and design, analysis and interpretation of data; as well as drafting the article and revising it critically for important intellectual content. Gave final approval.
Lisa M. Knowlton: substantial contributions to conception and design, analysis and interpretation of data; as well as drafting the article and revising it critically for important intellectual content. Gave final approval.
Dan E. Azagury: substantial contributions to conception and design, analysis and interpretation of data; as well as drafting the article and revising it critically for important intellectual content. Gave final approval.
Aussama K. Nassar: substantial contributions to conception and design, analysis and interpretation of data; as well as drafting the article and revising it critically for important intellectual content. Gave final approval.
James N. Lau: substantial contributions to conception and design, analysis and interpretation of data; as well as drafting the article and revising it critically for important intellectual content. Gave final approval.
Katherine D. Arnow: substantial contributions to conception and design, analysis and interpretation of data; as well as drafting the article and revising it critically for important intellectual content. Gave final approval.
Amber W. Trickey: substantial contributions to conception and design, analysis and interpretation of data; as well as drafting the article and revising it critically for important intellectual content. Gave final approval.
Carla M. Pugh: substantial contributions to conception and design, acquisition, analysis and interpretation of data; as well as drafting the article and revising it critically for important intellectual content. Gave final approval.

The authors report no conflicts of interest.
Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.
ISSN: 0003-4932/20/27203-0523
DOI: 10.1097/SLA.0000000000004207

Artificial intelligence (AI) is a scientific approach that uses theories and mathematical algorithms to give computer systems the ability to perform tasks that normally require human intelligence.[1,2] Major advances in processing speed, cloud storage, and the availability of annotated, big data have greatly facilitated the broad utilization of AI across a variety of industries. In healthcare, AI is projected to have a significant, positive impact on clinical workflows, patient outcomes, and accurate image interpretation.[3] Notable limitations of AI include bias, lack of transparency of factors associated with outcome, and variability in accuracy across systems and data types.

In the surgical profession, AI has gained significant popularity as a tool to analyze surgical videos. Industry and academic groups are both exploring ways to efficiently collect, store, and analyze surgical videos with the ultimate goal of providing decision support to improve surgical quality and patient outcomes.[4–7] At a recent Surgical Data Science conference, computer scientists, engineers, and surgeons agreed that the single, most significant barrier to wide utilization of AI in the surgical profession is access to large surgical datasets greater than 1000 entries.[8] Without access to data, it is difficult to develop and validate data analytic approaches, data processing standards, and user feedback.[8,9]

Laparoscopic cholecystectomy is an excellent procedure to explore the utility of AI as there are numerous peer-reviewed papers and reports that promote and recommend achieving the critical view of safety before dividing the cystic artery and duct.[10–12] If AI were found to be accurate and efficient in identifying when the critical view of safety was achieved, then this approach could be used for decision support, anatomical benchmarking of favorable and unfavorable anatomy, and for digital documentation of operative procedures.[5]

In 2019, our team partnered with a local industry group to review their previously acquired laparoscopic cholecystectomy video database with over 1000 entries to investigate the reliability and utility of AI-driven procedure segmentation and annotation. This study tests the hypothesis that AI accuracy in identifying the critical view of safety as well as other intraoperative events will be significantly correlated with disease severity. In addition, it is known that the critical view of safety is not always achieved; however, there is a paucity of studies that report incidence or causality. There are several

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

reasons why the critical view of safety may not be achieved including surgeon experience, technique or patient factors such as anatomical variations or disease severity. A secondary aim was to investigate the potential for AI to provide insight into the factors influencing variance in operative approach and outcomes, as well as to make inference regarding the potential contribution of AI as a quality assurance tool.

METHODS

Study Design

This study is a retrospective analysis of a multicenter laparoscopic cholecystectomy video database, collected from 2011 to 2019. The video database was obtained from a local company— Theator Global Surgery Index—General Surgery [Theator Inc, San Mateo, CA]. Videos of 1051 laparoscopic cholecystectomies were collected from 31 surgeons practicing at 2 institutions after approval from each site's Institutional Review Board.

AI Analysis of Video Database

Leveraging artificial intelligence (AI) algorithms were developed by Theator Inc, all videos underwent evaluation and annotation for surgical phases, intraoperative events, disease severity (Parkland scale), and fulfillment of critical view of safety (CVS) criteria.[13–16] The Parkland grading scale extends from grade 1 (normal appearing gallbladder) to grade 5 (necrosis, adhesions obscuring the gallbladder, or preoperative perforation). The AI system comprises 2 main modules: a temporal-ConvNet model and a temporal-sequence model (Fig. 1).

To maximize accuracy and balance computational effort, we used a standard, previously validated, video segmentation protocol. As described in Kay et al,[17] the surgical videos were segmented into short clips of 64 consecutive frames (images). Given the standard frame rate of 25 FPS, each video segment was 2.56 seconds long. Each segment (clip) was then processed with a 3D Deep Convolutional Neural Network (DCNN).[18] The 3D-DCNN was able to learn surgical context in both the spatial (surgical anatomy and operative field of view) and temporal (procedure steps and operative flow) domains. To further facilitate sequence and temporal modeling, a temporal Convolutional Network (temporal-ConvNet) was used to classify each second into an initial, suggested set of hypothesized annotations. The short segment annotations hypothesized by the temporal-ConvNet were then tested by a different temporal-sequence model, a Long Short-Term Memory network, that is capable of processing and making predictions on time series data with unknown duration between important events within the series.[19]

The special structure of the Long Short-Term Memory network allowed analysis of the predictions from the first model and attend to close-by or distant surgical events via a memory mechanism. For example, as surgery progresses there might be a short glimpse of a particular event and the model needs to ‘‘remember’’ this occurrence when reaching a decision junction of whether or not the event occurred in the ‘‘usual’’ expected sequence. The resulting output of the algorithm is shown in Figure 1. Each laparoscopic cholecystectomy procedure was then classified for Parkland Scale severity using AI evaluation.

Surgeon Review

Rater training for CVS recognition was accomplished by distribution of a single figure showing the 3 components of the CVS.[15] A representative sample of the 1051 videos was selected,

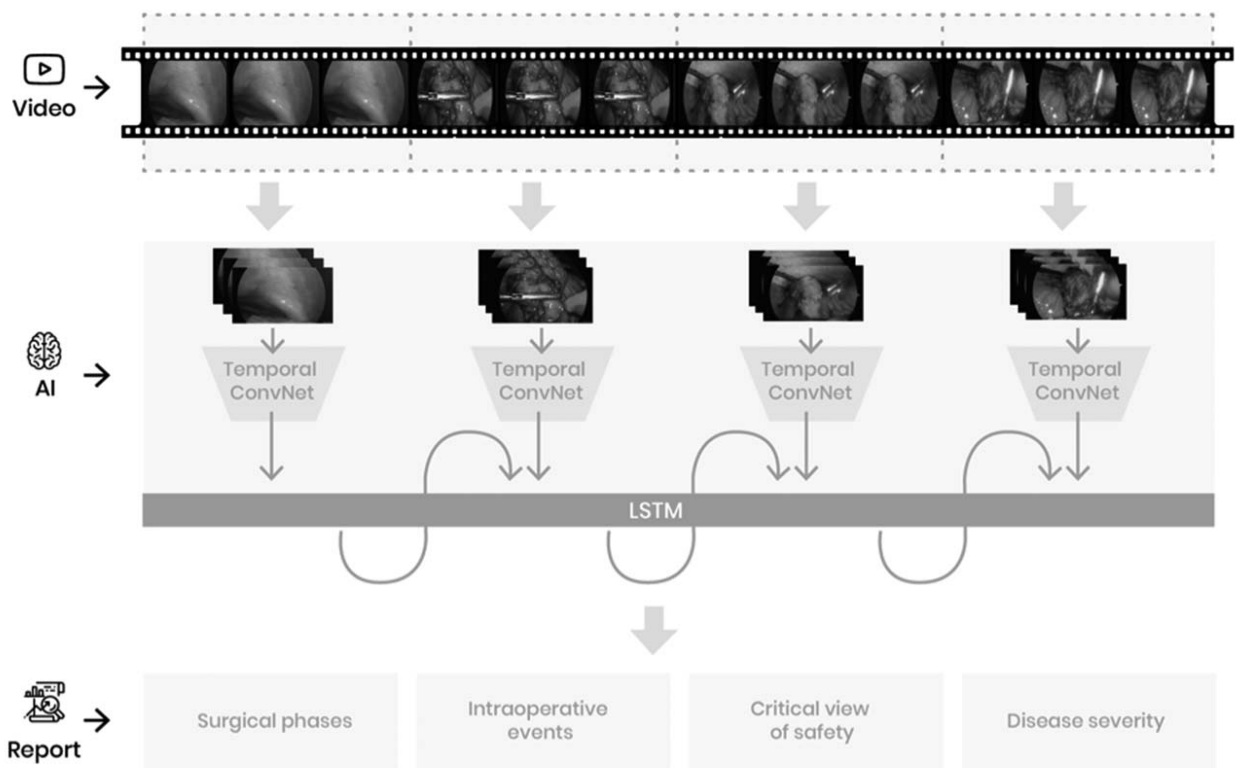

FIGURE 1. AI system description, mapping a surgical video to an intraoperative report. Video is segmented into short clips which are then fed to a Temporal-ConvNet. Each second is classified to produce a predefined set of annotations. The predictions are then sequentially processed by a temporal-sequence model (LSTM) to achieve the final surgical report annotations.

© 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

FIGURE 2. Surgeon video review interface in Theator software application.

randomly for the Parkland grade 1–2 (n = 50) and including all Parkland grade 3–5 with $\geq$1 intraoperative events (n = 335) that were annotated to contain an intraoperative event. Eight board-certified surgeons performed a focused review of the annotated sections showing CVS and intraoperative events using the Theator software interface (Fig. 2). Through the use of the interface, surgeons navigated directly to the step or event of interest and reviews were accomplished at a rate of approximately 50 videos per hour. Surgeons recorded whether AI annotations were correct for achievement of each CVS component and the Parkland Grade. When AI annotations were determined to be incorrect, the correct CVS and Parkland Grade information was provided by the surgeon rater.

Statistical Analysis

Interrater agreement between surgeons on achievement of CVS and each component was assessed with Cohen's kappa using a sample of 25 high-severity videos (Parkland 3–5). A subset of 10 videos were randomly selected within each surgeon's set of videos and were also scored by another surgeon—5 videos by 1 surgeon and 5 videos by another surgeon therefore, each surgeon had 10 of their videos overlap with another surgeon. To evaluate the validity of the AI annotation, AI-surgeon interrater agreement for annotation of CVS components was assessed with Cohen's kappa, calculated overall and by severity level (Parkland 1–2 vs 3–5). Differences between agreement rates in low and high severity cases were assessed with chi-square tests.

Achievement of CVS measured as complete (all components) as well as individual CVS components were compared between low- and high-severity cases with chi-square tests. Intraoperative event associations with lack of CVS achievement and severity were examined using ordinal logistic regression. Statistical significance was assessed at the level of $P < 0.05$. Statistical analyses were performed using Stata v. 15.1 (StataCorp LLC, College Station, TX).

RESULTS

Interrater Reliability

AI annotation was compared to surgeon review to determine the reliability of identification of CVS components and intraoperative events. AI-surgeon agreement for all CVS components exceeded 81% (kappa 0.44–0.62) (Table 1). A subset analysis for

TABLE 1. Kappa Values and Percent Agreement Between AI and Surgeon Ratings for the Critical View of Safety (All Three) and Each Component for All Cases and by Disease Severity

	All Cases (n = 385)		Low-severity Cases Grades 1–2 (n = 50)		High-severity Cases Grades 3–5 (n = 335)		*P* Value
CVS Components	Kappa	% Agree	Kappa	% Agree	Kappa	% Agree	% Agree
Hepatocystic Triangle	0.56	81	0.53	76	0.55	82	0.30
Cystic plate	0.62	84	0.57	80	0.63	84	0.42
2 structures	0.44	81	0.17	56	0.51	85	<0.001
All 3 components	0.56	93	0.31	80	0.64	95	<0.001

© 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 2. Operative Achievement of the CVS (*All Three Components*) and Individual CVS Components for All Cases Compared to Low and High Severity Cases

	% Achieved			
CVS Component	**All Cases**	**Low-severity Cases (Grades 1, 2)**	**High-severity Cases (Grades 3, 4, 5)**	***P* Value**
Hepatocystic Triangle	37%	41%	34%	0.031
Cystic plate	38%	43%	36%	0.027
2 structures	14%	16%	13%	0.16
All 3 components	9%	10%	9%	0.66

low and high severity cases was performed. AI-surgeon agreement was significantly higher in greater severity cases for the presence of all 3 components and for 2-structures compared to low severity ($P < 0.001$). No significant difference in agreement was found for cystic plate or hepatocystic triangle components of CVS. Surgeon-to-surgeon agreement was 100% for all 3 components present (k 1.0), 92% for 2 structures (k.63), and hepatocystic triangle (k.80), and 96% for cystic plate (k.90).

AI Annotation

Only 9% of all 1051 videos annotated by AI revealed identification of all 3 CVS components. Of the individual components, 2 structures were the least commonly noted, identified in only 14% of cases. Hepatocystic triangle (41% vs 34%, $P = 0.031$) and cystic plate (43% vs 36%, $P = 0.027$) were more frequently identified in low-severity cases than high-severity cases. No significant differences were found between identification of 2-structures or all 3 components between low- and high-severity cases (Table 2).

AI annotated 836 intraoperative events. High severity cases had higher mean event counts than low severity cases (0.98 vs 0.40 events/case, $P < 0.001$). We found that increased severity was associated with increased odds of unplanned events such as gallbladder gallstone spillage (OR 3.1 [95% CI 2.1,4.6], $P < 0.001$) and gallbladder bile leak (OR 1.5 [95% CI 1.2,1.9], $P < 0.001$), as well as surgeon planned events such as gallbladder decompression (OR 2.1 [95% CI 1.5, 3.0], $P < 0.001$) and drain insertion (OR 7.7 [95% CI 5.5, 10.8], $P < 0.001$). Except for gallstone spillage (OR 2.1 [95% CI 1.3, 3.2], $P < 0.001$), the number of CVS components not achieved was not significantly associated with intraoperative events (Table 3).

DISCUSSION

This study investigated the reliability and clinical utility of AI annotation of surgical videos. Regarding utility, one of our key findings was that AI annotation and segmentation of surgical videos allows for extremely efficient video review (~ 50 surgical videos in 1 h). This finding has vast implications regarding surgical education, coaching, and quality oversight. In addition, AI annotation may facilitate novel investigations and discoveries revealing the complex nature of the surgical process and the numerous, moment-to-moment effectors that impact surgical decision making.[20–24] While it has been shown, in theory, that a wide variety of factors affect surgical decision making, our ability to systematically research these factors has been limited.[25] Efficiency of video review has been a major barrier for utilization of videos as a data source for empirical investigation. Our study found AI annotation may help to lessen barriers and allow for new research that may help to close the information gap between surgical processes and surgical patient outcomes.

Reliability of AI

When evaluating the reliability of AI annotation, we found that there were high levels of agreement between AI and surgeons when identifying the components of the critical view of safety. The percent agreement for all 3 CVS components combined across all cases was 93% and had a range of 80% to 95% when stratifying the cases by disease severity. The lowest percent agreement was for low severity cases in which AI-Surgeon agreement for identification of the *2 structures* component of the CVS was 56%. This discrepancy serves as a reminder to the surgical community that there may be specific and important areas within an AI algorithm that do not align with surgical decision-making and practice.[4] Ultimately, these differences may serve to advance the science of surgical process research and further enhance our utilization of AI as a tool that can facilitate systematic investigations into surgical decision-making and technique.[5] Conversely, it is also possible that AI will continue to discover a small but persistent number of nuanced events that are simply difficult to explain. Overall, this finding supports our hypothesis that disease severity will have an effect on AI accuracy in identifying the critical view of safety as well as other intraoperative events.

TABLE 3. Relationship Between Intraoperative Events and Disease Severity As Well As Intraoperative Events and CVS Components Not Achieved

Ordinal Logistic Regression			**Disease Severity**			**# CVS Components Not Achieved**		
Intraoperative Events	**N**	**%**	**OR**	**95% CI**	***P***	**OR**	**95% CI**	***P***
Drain insertion	154	14.7	7.7	5.5–10.8	<0.001	1.3	0.9–1.8	0.18
Suspected bowel injury	5	0.5	4.6	1.0–22.2	0.06	0.9	0.1–5.4	0.90
Gallbladder gallstone spillage	100	9.5	3.1	2.1–4.6	<0.001	2.1	1.3–3.2	<0.001
Gallstone extraction	18	1.7	2.8	1.3–6.1	0.01	1.0	0.4–2.2	0.90
Use of electrocautery adjacent to colon	11	1.1	2.2	0.8–6.0	0.12	3.2	0.7–14.9	0.14
Gallbladder decompression	97	9.2	2.1	1.5–3.0	<0.001	1.2	0.8–1.8	0.50
Cystic duct bile leak	23	2.2	1.9	0.9–3.9	0.07	1.3	0.6–2.9	0.46
Gallbladder sludge spillage	45	4.3	1.6	0.9–2.7	0.10	1.5	0.8–2.6	0.22
Gallbladder bile leak	363	35.5	1.5	1.2–1.9	0.001	1.2	0.9–1.6	0.13
Cystic duct gallstone spillage	15	1.4	1.2	0.5–2.9	0.63	3.3	1.0–11.7	0.06
Cholangiogram	5	0.5	0.4	0.1–1.9	0.25	0.4	0.1–2.2	0.32

© 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

AI as a Quality Assurance Tool

Another key finding when reviewing the 1051 laparoscopic cholecystectomy videos was that only 9% of cases achieved CVS, an evidence-based recommendation. This was an unexpected finding. While there are a paucity of studies systematically investigating and indicating the true incidence of achieving the CVS, our team did not expect such a low percentage. What is also noteworthy is that there were no common bile duct injuries in our database. This finding in no way indicates that the CVS method is not useful. However, it does raise a number of questions regarding utilization and adherence to evidence-based approaches as well as the extent to which practitioners use the CVS recommendations as a guideline but do not adhere to the strict anatomical definition and implementation of all 3 CVS components. While the answers to these questions are out of the scope of this current study, this is an example of the complexity and nuance of surgical decision-making. Anecdotally, during the surgeon video reviews, it was noted, on more than 1 occasion, that arterial anomalies prevented achievement of the *2-structures* CVS component prior to the first clip being applied; however, the surgeon achieved the cystic plate CVS component prior to clipping the cystic duct. In other instances, there was a clear opportunity to achieve textbook CVS and it simply was not done. Likewise, some of the cases were notably difficult and a number of intraoperative events ensued that changed the course of dissection, anatomical identification, and clip placement. Many of the other findings were firmly in line with what was to be expected. Disease severity had a significant effect on achieving 2 of the CVS components (hepatocystic triangle and cystic plate). In addition, intraoperative events occurred more frequently in high-severity cases. Moreover, the lower the number of CVS components achieved, the higher the likelihood of gallstone spillage. Overall, this level of quality review and oversight was made possible due to AI annotation of CVS components and intraoperative events.

Study limitations relate to the use of a single-institution analysis of an industry-procured multiinstitutional database and utilization of AI annotations generated by an industry group. The single-institution analysis creates local bias in the surgeon training and video review results. For broader generalizability additional institutions would need to analyze the same set of videos. Despite this limitation, our team was able to achieve high levels of AI–Surgeon agreement with minimal surgeon training which may point to the ease of anatomical inspection and clarity of the recommended guidelines for CVS components. Comparison of AI analysis and patient outcomes would have been ideal but was not possible with the current data set. Even though patient outcomes were collected, the combination of low attainment of CVS and low occurence rate of adverse patient outcomes prevented any meaningful inference. With a much larger data set than the current 1051 videos and AI annotation, such an anylysis may be possible in the future. Regarding the industry-based AI annotations, our team's review of CVS components revealed some differences in AI interpretation of the *2 structures* CVS component which indicates that more research or a consensus agreement may be necessary for AI to fully capture and annotate this component in the context of wide variations in dissection strategies and endpoints. In other words, how much of the artery and duct need to be visualized before everyone agrees that there are only *2* structures going to the gallbladder. This is not perfectly defined in the CVS literature nor has it been empirically tested from a patient safety standpoint.[26] Another weakness of the reviewing only the pre-existing, industry-based AI annotation was the potential for AI missing other intraoperative events, for example, the lack of annotation for intraoperative bleeding. When the company began their database annotation there was no generalizable rating system for surgical bleeding. Recently, a rating system was published and will greatly facilitate future work and annotations of this database.[27] Despite this omission, we were able to achieve a clinically valid review of the 1051 cases and draw meaning from the relationships between AI-Surgeon CVS annotation, the effect of disease severity on CVS components and intraoperative events.

Future work will involve continued refinement of AI annotation strategies and advances in surgeon involvement in the annotation process. It has been noted that computer scientists and engineers are calling for closer partnerships with surgeons and other practitioners in order to advance the utility of AI for surgical care processes. As such, if the goal of real-time decision support and improvement in quality is to be achieved, the process of collecting, storing, and accessing large databases must be realized and streamlined in addition to getting surgeons engaged in the annotation process.

CONCLUSION

AI annotation allows for efficient video review and is a promising quality assurance tool. Disease severity has a significant impact on its use and surgeon oversight is still required to interpret the results. Continued refinement may improve AI applicability and allow for automated assessment.

ACKNOWLEDGMENTS

The authors thank Kenneth Perrone, MD, Anna Witt, Cassidi Goll, Brett Wise, Su Yang, Hossein Mohamadipinah, PhD, Tamir Wolf, MD, PhD, Theator Inc.

REFERENCES

1. McCarthy J, Hayes PJ. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In: *Readings in Artificial Intelligence*. Amsterdam, Netherlands: Elsevier; 431–450.
2. Bellman R. An Introduction To Artificial Intelligence Available at: http://archive.org/details/AnIntroductionToArtificialIntelligence. 1978. Accessed March 31, 2020.
3. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44–56.
4. Hashimoto DA, Rosman G, Rus D, et al. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018;268:70–76.
5. Mascagni P, Fiorillo C, Urade T, et al. Formalizing video documentation of the Critical View of Safety in laparoscopic cholecystectomy: a step towards artificial intelligence assistance to improve surgical safety. *Surg Endosc*. 2019;34. 2709–2714.
6. Samareh A, Chang X, Lober WB, et al. Artificial intelligence methods for surgical site infection: impacts on detection, monitoring, and decision making. *Surg Infect (Larchmt)*. 2019;20:546–554.
7. Kapoor R, Walters SP, Al-Aswad LA. The current state of artificial intelligence in ophthalmology. *Surv Ophthalmol*. 2019;64:233–240.
8. Maier-Hein L, Vedula SS, Speidel S, et al. Surgical data science for next-generation interventions. *Nat Biomed Eng*. 2017;1:691–696.
9. Dzobo K, Adotey S, Thomford NE, et al. Integrating artificial and human intelligence: a partnership for responsible innovation in biomedical engineering and medicine. *OMICS*. 2019;24. 247–263.
10. Strasberg SM, Hertl M, Soper NJ. An analysis of the problem of biliary injury during laparoscopic cholecystectomy. *J Am Coll Surg*. 1995;180:101–125.
11. Strasberg SM. Biliary injury in laparoscopic surgery: part 1. Processes used in determination of standard of care in misidentification injuries. *J Am Coll Surg*. 2005;201:598–603.
12. Strasberg SM. Biliary injury in laparoscopic surgery: part 2. Changing the culture of cholecystectomy. *J Am Coll Surg*. 2005;201:604–611.
13. Lee W, Jang JY, Cho J-K, et al. Does surgical difficulty relate to severity of acute cholecystitis? Validation of the parkland grading scale based on intraoperative findings. *Am J Surg*. 2018;219. 637–641.
14. Madni TD, Nakonezny PA, Barrios E, et al. Prospective validation of the Parkland Grading Scale for Cholecystitis. *Am J Surg*. 2019;217:90–97.
15. 087686-180213-Critical-View-of-Safety.pdf Available at: https://jnjinstitute.com/sites/default/files/2019-05/087686-180213-Critical-View-of-Safety.pdf. Accessed March 31, 2020.

© 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.

16. Laparoscopic Cholecystectomy Educational Tutorial: Explaining the Critical View of Safety on Vimeo Available at: https://vimeo.com/190486987. Accessed March 31, 2020.
17. Kay W, Carreira J, Simonyan K, et al. The Kinetics Human Action Video Dataset. *arXiv:170506950 [cs]*. Available at: http://arxiv.org/abs/1705.06950. 2017. Accessed April 3, 2020.
18. Carreira J, Zisserman A. Quo vadis, action recognition? a new model and the kinetics dataset. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017 (pp. 6299–6308). Las Alamitos, CA: IEEE.
19. Deep learning | Nature Available at: https://www.nature.com/articles/nature14539. Accessed March 31, 2020.
20. Moulton CE, Regehr G, Mylopoulos M, et al. Slowing down when you should: a new model of expert judgment. *Acad Med*. 2007;82:S109–116.
21. Flin R, Youngson G, Yule S. How do surgeons make intraoperative decisions? *Qual Saf Health Care*. 2007;16:235–239.
22. Patel VL, Arocha JF, Kaufman DR. A primer on aspects of cognition for medical informatics. *J Am Med Inform Assoc*. 2001;8:324–343.
23. Orasanu J, Connolly T. The reinvention of decision making. In: *Decision Making in Action: Models and Methods*. Westport, CT: Ablex Publishing; 1993:3–20.
24. Klein G. Naturalistic decision making. *Hum Factors*. 2008;50:456–460.
25. Pugh CM, Santacaterina S, DaRosa DA, et al. Intra-operative decision making: more than meets the eye. *J Biomed Inform*. 2011;44:486–496.
26. Deal SB, Alseidi AA. Concerns of quality and safety in public domain surgical education videos: an assessment of the critical view of safety in frequently used laparoscopic cholecystectomy videos. *J Am Coll Surg*. 2017;225:725–730.
27. Lewis KM, Li Q, Jones DS, et al. Development and validation of an intraoperative bleeding severity scale for use in clinical studies of hemostatic agents. *Surgery*. 2017;161:771–781.

© 2020 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.