

Review

Open Access



Artificial intelligence in hepatopancreaticobiliary surgery - promises and perils

Christina Boutros¹, Vivek Singh², Lee Ocuin¹, Jeffrey M. Marks¹, Daniel A. Hashimoto³

¹Department of Surgery, University Hospitals Cleveland Medical Center, Cleveland, OH 44106, USA.

²Boston University School of Medicine, Boston, MA 02215, USA.

³Department of Surgery, Hospital of the University of Pennsylvania, University of Pennsylvania Perelman School of Medicine, PA 19104, USA.

Correspondence to: Dr. Daniel Hashimoto, Department of Surgery, Hospital of the University of Pennsylvania, University of Pennsylvania Perelman School of Medicine, 4 Silverstein Pavilion, PA 19104, USA. E-mail: daniel.hashimoto@pennmedicine.upenn.edu

How to cite this article: Boutros C, Singh V, Ocuin L, Marks JM, Hashimoto DA. Artificial intelligence in hepatopancreaticobiliary surgery - promises and perils. *Art Int Surg* 2022;2:213-23. <https://dx.doi.org/10.20517/ais.2022.32>

Received: 4 Oct 2022 **First Decision:** 17 Nov 2022 **Revised:** 15 Dec 2022 **Accepted:** 28 Dec 2022 **Published:** 30 Dec 2022

Academic Editors: Andrew A. Gumbs, Derek O'Reilly **Copy Editor:** Ying Han **Production Editor:** Ying Han

Abstract

Research and development in artificial intelligence (AI) has been experiencing a resurgence over the past decade. The rapid growth and evolution of AI approaches can leave one feeling overwhelmed and confused about how these technologies will impact hepatopancreaticobiliary (HPB) surgery, the obstacles to its clinical translation, and the role that HPB surgeons can play in accelerating AI's development and ultimate clinical impact. This review outlines some of the basic terminology and current approaches in surgical AI, obstacles to further development and translation of AI, and how HPB surgeons can influence its future in surgery.

Keywords: Artificial intelligence, computer vision, natural language processing, machine learning, hepatopancreaticobiliary surgery

INTRODUCTION

Research and development in artificial intelligence (AI) has been experiencing a resurgence over the past decade. Historically, AI was first coined in 1956 in the Dartmouth Summer Research Program, a workshop organized and attended by scientists who are now recognized as the founders of AI as a field of study. With



© 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.



roots in mathematics, statistics, neuroscience, linguistics, and even philosophy, AI has evolved from programs that took advantage of rule-based systems based on expert knowledge to automated feature extraction through deep learning^[1]. The rapid growth and evolution of AI approaches can leave one feeling overwhelmed and confused about how these technologies will impact hepatopancreaticobiliary (HPB) surgery, the obstacles to its clinical translation, and the role that HPB surgeons can play in accelerating AI's development and ultimate clinical impact. Thus, we review some of the basic terminology and current approaches in surgical AI and how HPB surgeons can influence its future in surgery.

MACHINE LEARNING

Maier-Hein *et al.* define surgical data science (SDS) as the capture, organization, analysis, and modeling of data sets with the aim of improving interventional healthcare^[2]. AI, more specifically, machine learning, has played a significant role in the advancement of surgical data science, though it is important to note that machine learning is not the only technique used in SDS. To better understand AI work that has been performed thus far in HPB surgery, it is important to recognize that there are several learning strategies within machine learning. Two popular approaches are classical machine learning, where humans specifically guide algorithms to include specific features in their analyses, and neural networks, where features are automatically derived by the algorithm. Neural networks are designed to function in a manner akin to biological neural networks - a series of computational units (i.e., neurons) are arranged in layers and process input data to yield a target output. Deep learning refers to neural networks that are arranged in greater than 3 layers (i.e., input layer, > 1 hidden layer, output layer). The use of many hidden layers enables the network to learn much more complex relationships than would be possible through a simple 3-layer neural network. Each neuron within the network receives input data, processes it using a mathematical function, and yields an output. Each neuron may be weighted differently depending on the overall performance of the network, and these weights are tuned over multiple examples to generate optimal performance on a given task. Deep learning is a technique that uses layers of computation to extract information from an input in progressive fashion until a desired output is produced^[3]. Progressive extraction refers to different levels of information. For example, initial layers in a deep learning network may extract such things as edges in an image, while subsequent layers may utilize information on those edges to identify faces and then specific people.

Common learning strategies that clinicians are likely to encounter in the AI literature are supervised, unsupervised, and semi-supervised learning. Supervised learning refers to task-driven learning wherein an algorithm learns from data that has been annotated by humans^[4]. For example, a researcher may have annotated a dataset of the critical view of safety in hundreds of photographs from laparoscopic cholecystectomy. An algorithm would "learn" from these photographs and annotations and subsequently attempt to label new images from laparoscopic cholecystectomies with the critical view of safety. While supervised learning through neural networks removes the effort needed to select features manually for analysis, it has several challenges. Large amounts of data are typically needed to optimize the performance of neural networks, especially deep learning. Furthermore, the data must be annotated to appropriately identify the target output. That is, if one wishes to accurately identify photos of the critical view of safety in cholecystectomy, a dataset composed primarily of neurosurgical photographs may not have much value. In supervised learning, the labels on which the algorithms are trained are a critical part of the data.

Unsupervised learning, on the other hand, does not rely on labels to learn phenomena. Instead, unsupervised learning analyzes the patterns inherent in the data to determine whether identifiable groups are present. Stated perhaps more simply, unsupervised learning allows the data to be clustered into groups with similar features, and a human can interpret those clusters to determine whether features may identify

specific groups, such as patients most likely to experience a complication.

Semi-supervised learning combines elements of both supervised and unsupervised learning. A small amount of labeled data can be combined with unlabeled data to achieve weakly supervised analysis of surgical data. For example, one application of semi-supervised learning is to use a small amount of labeled data to train an algorithm and to augment algorithmic performance with the unlabeled data. For example, unlabeled cholecystectomy videos have been used to improve the performance of an algorithm that has been trained on a small number of labeled videos of cholecystectomy^[5].

APPLICATIONS OF ARTIFICIAL INTELLIGENCE IN HEPATOPANCREATICOBILIARY SURGERY

While AI and machine learning have great success in other fields of medicine, their use in surgery has just started to increase. More manuscripts that incorporate AI and ML principles to advance the practice of surgery, and specifically HPB surgery, are being published. Bektas *et al.* have recently reviewed applications of AI in HPB surgery^[6]. Three important areas of application to HPB surgery that clinicians should familiarize themselves with include machine learning for tabular data, natural language processing, and computer vision [Figure 1].

Machine learning for tabular data

Tabular data simply refers to data that can be organized into rows and columns and is perhaps the type of data with which clinicians are most familiar. Tabular data is utilized in data collected from claims records, electronic medical records, or large national datasets. The performance of machine learning on such data is variable and the literature has been quite heterogeneous regarding the state-of-the-art, with manuscripts suggesting either very poor or impossibly good performance of predictive algorithms^[7].

Common sources of tabular data include the electronic medical record of a given institution, claims databases such as the Nationwide Inpatient Sample, or specific registries such as the National Cancer Database [Table 1]. Specific registries can take into account clinical variables of interest for given procedures. For example, Merath *et al.* utilized the American College of Surgeons (ACS) National Surgical Quality Improvement Program to identify patients undergoing liver, pancreatic and colorectal surgery from 2014 to 2016^[8]. Decision tree models were utilized to predict the occurrence of a broad range of complications, including stroke, wound dehiscence, cardiac arrest, and progressive renal failure, with accuracy outperforming known risk stratification tools like the ASA and ACS surgical risk calculator^[8]. Machine learning has been used to create algorithms to predict specific issues such as postoperative diabetes mellitus after partial pancreatectomy. All patients undergoing partial pancreatectomy in a single institution were analyzed from 2015 to 2019. Machine learning was utilized to predict the development of postoperative diabetes mellitus after one year. Their algorithm was able to accurately predict the development of diabetes with 87% and 85% sensitivity and specificity, respectively^[9].

Natural language processing

Natural language processing (NLP) is a field that focuses on using automated methods to organize and make sense of language - both spoken and written. Clinically, NLP algorithms are especially useful for interpreting and structuring text from electronic health records (EHR). Historically, vocabulary- and rule-based NLP approaches have relied on lists of words or phrases with multiple variations of a phrase or word. This, however, becomes untenable as clinical scenarios become more complex. Furthermore, such approaches require more predictable formats of free text (e.g., structured notes) to optimize performance. For example, Al-Haddad *et al.* utilized natural language processing to create a registry of patients with

Table 1. Summary of studies showcasing representative applications of AI in HPB surgery

Study	Author	Year	Target	Tool	Data	Conclusion
Use of Machine Learning for Prediction of Patient Risk of Postoperative Complications After Liver, Pancreatic, and Colorectal Surgery	Merath <i>et al.</i> ^[9]	2020	Liver, pancreatic, and colorectal surgery	Decision trees	ACS NSQIP	Decision tree models were utilized to predict the occurrence of a broad range of complications, outperforming known risk stratification tools like the ASA and ACS surgical risk calculator
Natural language processing for the development of a clinical registry: a validation study in intraductal papillary mucinous neoplasms	Al-Haddad <i>et al.</i> ^[11]	2010	IPMN surveillance	Natural Language Processing	Single institution medical records	Regenstrief EXtraction Tool (REX) used to extract pancreatic cyst patient data. Detected patients with IPMN with high sensitivity
Automated pancreatic cyst screening using natural language processing: a new tool in the early detection of pancreatic cancer	Roch <i>et al.</i> ^[12]	2015	Pancreatic cyst surveillance	Vocabulary- and rule-based NLP	Single-institution medical records	Key words and phrases searched within the electronic medical record to identify patients with pancreatic cysts with high sensitivity and specificity to build a registry for patients at risk of pancreatic cancer
A computer vision platform to automatically locate critical events in surgical videos: Documenting safety in laparoscopic cholecystectomy	Mascagni <i>et al.</i> ^[19]	2021	Laparoscopic Cholecystectomy	Deep learning and rule-based computer vision	Cholecystectomy videos	Successfully isolated a short segment video clip in which the critical view of safety was obtained from videos
Artificial intelligence for intraoperative guidance	Madani <i>et al.</i> ^[21]	2022	Laparoscopic Cholecystectomy	Deep learning, computer vision	Cholecystectomy videos	Deep learning model trained on expert annotations can accurately highlight safe/unsafe dissection areas
Artificial intelligence prediction of cholecystectomy operative course from automated identification of gallbladder inflammation	Ward <i>et al.</i> ^[22]	2022	Laparoscopic Cholecystectomy	Computer vision and Bayesian Models	Cholecystectomy videos	Automated identification of Parkland Grading Scale (PGS) used to predict the intraoperative course and likelihood of spilling bile

intraductal papillary mucinous neoplasms (IPMN) in their health system using the Regenstrief EXtraction Tool (REX) to extract pancreatic cyst patient data from medical text files. Their program was able to detect patients with IPMN with high sensitivity and suggested that this was a potentially useful and reliable tool to identify patients with pancreatic cysts who require follow-up^[10]. Roch *et al.* performed a similar experiment, utilizing vocabulary- and rule-based NLP to create a registry of patients with pancreatic cysts in their hospital system. Key words and phrases were given to the program, which searched the electronic medical record and was able to identify patients with pancreatic cysts with high sensitivity and specificity. Their system helped capture patients with a risk of pancreatic cancer in a registry which can be utilized to monitor patients and aid in follow-up^[11].

More modern approaches to NLP have utilized deep learning techniques to minimize the amount of feature engineering required for good performance and to maximize performance on more natural forms of human language that require less structure and fewer explicit examples and rules. Advanced AI models such as Generative Pre-Trained Transformer 3 (GPT-3) are also able to generate human-like text to create de novo conversations and even works of literature^[12]. Because much of the text generated in medical encounters is semi-structured (e.g., History and Physicals, SOAP progress notes), state-of-the-art generative models may not be a necessity for simple NLP tasks in HPB surgery that facilitate billing and data extraction. However, improved NLP models can facilitate

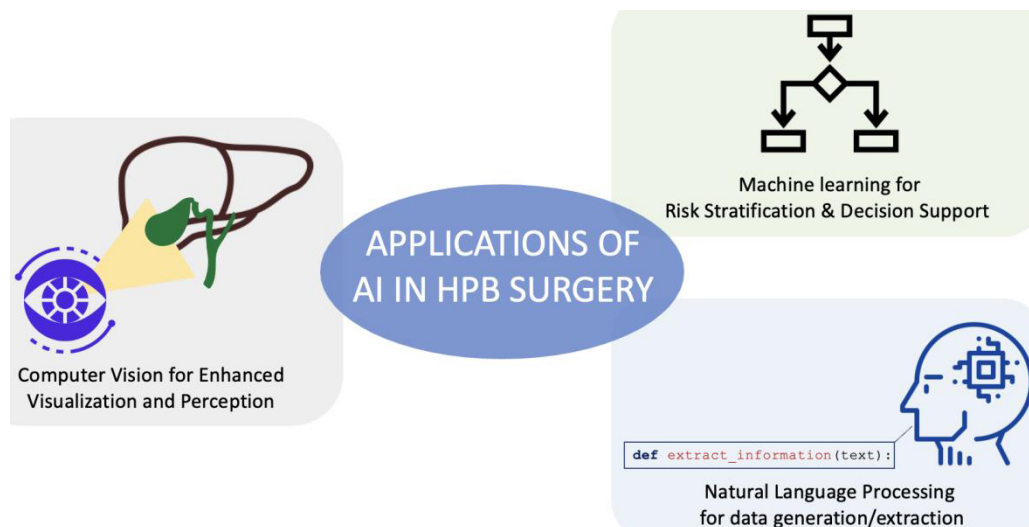


Figure 1. Overview of applications of AI in HPB surgery.

the advancement of voice recognition models that can increase the efficiency of patient encounters through automated documentation or decision support^[13].

Computer vision

Computer vision, the field concerning itself with machine understanding of visual inputs, has been a useful tool in the advancement of AI applications in HPB surgery. With applications ranging from identification of operative steps and instruments on the field to automated labeling of operative steps in surgical videos, recent approaches to computer vision in surgery have most commonly utilized deep learning, more specifically convolutional neural networks (CNN), to achieve its aims^[14]. CNNs use filters to, in effect, summarize the information contained in a given part of the data. For example, a data input that is 4×4 units can be processed by a filter that passes over the image as a 2×2 block that slides over the data and extracts weights [Figure 2]. Multiple convolutional layers can be utilized together to process complex data^[1].

Kenngott *et al.* in 2014 explored an “augmented reality” system that used robotic C arm cone beam computed tomography (CT) to help identify hepatic tumors and aid in achieving oncologic resection during laparoscopic liver surgery. By combining fiducial markers with CT images, they demonstrated precise identification of the location of hepatocellular carcinoma that could be used to help guide resection^[15]. Although Kenngott’s study did not utilize computer vision, there are certainly applications of computer vision in this realm. As computer vision technology and physics modeling of soft-tissue deformations continue to improve, techniques that do not require the use of fiducial markers have been developed to further assist in the intraoperative utilization of axial imaging overlays to augment real-time surgical planning^[16,17].

Perhaps the most investigated application of computer vision in HPB surgery is laparoscopic cholecystectomy. Mascagni *et al.* (2021) created a program, EndoDigest, which could identify critical events in surgical videos and generate clips documenting acquisition of the critical view of safety during a laparoscopic cholecystectomy^[18]. By identifying relevant portions of the case from video recordings, EndoDigest was able to successfully isolate a short segment video clip in which the critical view of safety was obtained in 91% of the test videos in the dataset^[18]. Such automatic indexing of surgical procedures can play an important role in improving the efficiency of video review and also enables other algorithms to be

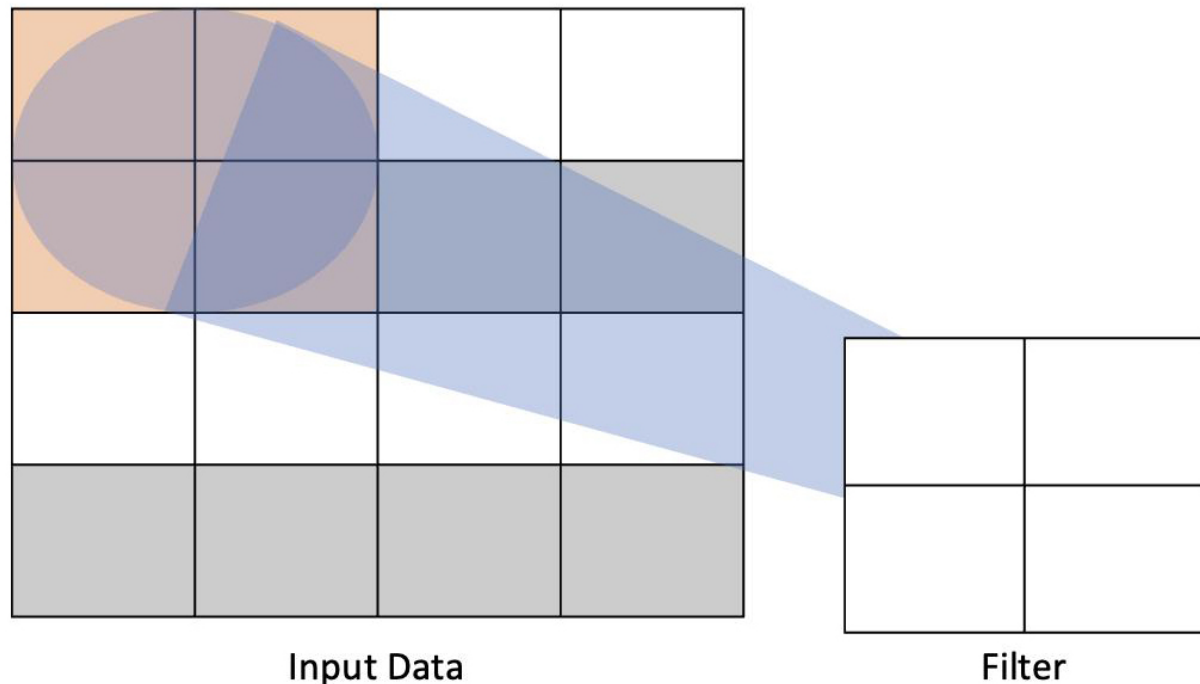


Figure 2. Example of how one layer of a convolutional neural network processes input data. A 4×4 unit input is processed by a 2×2 filter to extract relevant weights for the task according to the architecture of the network.

selectively activated to help with other tasks. For example, CVSnet, another algorithm from Mascagni *et al.* (2022), is an automated method to highlight whether or not a surgeon has achieved the three key components of the critical view of safety^[19]. Furthermore, Madani *et al.* (2022) created a deep learning computer vision model, trained on the annotations of experienced surgeons, to identify safe and unsafe zones of dissection in the hepatocystic triangle during a laparoscopic cholecystectomy with the aim of preventing injury due to visual misperception in the identification of anatomy^[20]. Such examples demonstrate that AI could be used to identify anatomy, with the goal of one day enabling real-time feedback in the operating room using a constellation of algorithms to aid surgeons and prevent adverse events and injuries.

Beyond intraoperative decision support at the level of identifying structures or safe and unsafe areas of dissection, computer vision can help to extract insights about patient characteristics. Ward *et al.* (2022) used a combination of computer vision and Bayesian models to identify the Parkland Grading Scale (PGS), a measure of gallbladder inflammation, and predict the intraoperative course of a laparoscopic cholecystectomy^[21]. In this manner, the authors could investigate whether increasing PGS score was associated with differences in the length of operative time, likelihood of attaining the critical view of safety, or likelihood of injuring the gallbladder and spilling bile. The computer vision model was able to accurately identify the PGS, and the work otherwise demonstrated that different surgeons were affected to different but systematic lengths by the level of inflammation of the gallbladder^[21].

OBSTACLES TO AI IN HPB SURGERY

The sheer volume of publications and discussion on AI in surgery would suggest that these technologies may be ready for primetime and are an inevitability in clinical practice. The reality, however, is that significant obstacles remain that prevent the translation of AI research into clinically usable and meaningful

applications. These obstacles exist across stakeholders, with clinicians, computer scientists, patients, hospitals, and even industry encountering issues that have prevented further progress in the field [Figure 3].

Lack of data

In the realm of computer vision research, data are a vital and often scarce currency. For algorithms to function, they must be fed a considerable quantity of training data before they are able to generate conclusions “independently”. However, the success or failure of a given project depends on not only an abundance of data but an abundance of the right kind of data. While simple in premise, the capture, processing, and storage of data can represent a herculean task, especially in the realm of computer vision. Firstly, institutional limitations on the sharing and acquisition of medical media have created necessary but ample logistical hurdles in the generation of new datasets^[1]. Likewise, inter-institutional variability in image quality, file format, and capture process has effectively quashed the consolidation of multiple datasets into a single, more robust unit. Finally, the storage and processing of these data requires significant resources - not by way of computing power, but manpower. Datasets must be carefully tracked, evaluated, and managed depending on their use for a given project. These problems are compounded in the field of surgery, where raw data often represent video footage of procedures. Unlike still images, surgical videos generate complex spatiotemporal inputs that further exacerbate these issues.

Lack of ground truth

Once visual data have been appropriately captured, processed, and stored, they can begin to be used in the development of algorithms. However, at this stage, datasets lack the context needed to generate conclusions independently. That is, datasets are missing labels of objects, scenes, and events. The process of labeling, or annotation, is carried out by humans and serves to train an algorithm on how to interpret the raw visual data with which it is presented (e.g., supervised learning). Crucial to this premise is the establishment of a fundamental basis for data interpretation, or ground truth.

In theory, annotators establish ground truth by defining parameters such as tools, anatomical structures, and clinically significant events^[22]. In surgical applications, this process is limited by the degree of expertise required to make such annotations. Besides surgeons or surgical trainees, most annotators would require training on how to interpret the visual data presented from surgical procedures. For applications lacking a standardized operative procedure, like many in the HPB space, this process becomes even more difficult. Similarly, the lack of a standard operative procedure creates a wider margin for the interpretation of clinical events, which translates to greater annotator-annotator variability, even amongst experts^[23]

Explainability

The nature of so-called black box algorithms is a central concern of many ethicists when evaluating the moral permissibility of artificial intelligence. Deep learning systems are trained to recognize patterns from many thousands of data points but also to continuously adapt to novel data and modulate analyses accordingly^[24]. The resultant algorithms are highly sophisticated and at times highly accurate; however, this complexity comes at the expense of comprehension. While it is entirely possible to describe the annotation criteria or mathematical underpinnings of these algorithms, the self-learning aspect of black box algorithms generates outputs that are impossible for humans to understand.

The apparent lack of transparency in black box algorithms could serve to undermine trust in physicians and complicate the process of shared decision making^[25]. Moreover, the principle of informed consent would be nearly impossible to uphold. After all, if the process of a medical decision lacks a basis for explanation, educating patients on the risks and benefits of one outcome versus another would be a moot, if not impossible, point. In the field of surgery, this problem is further compounded by the acuity of the domain.

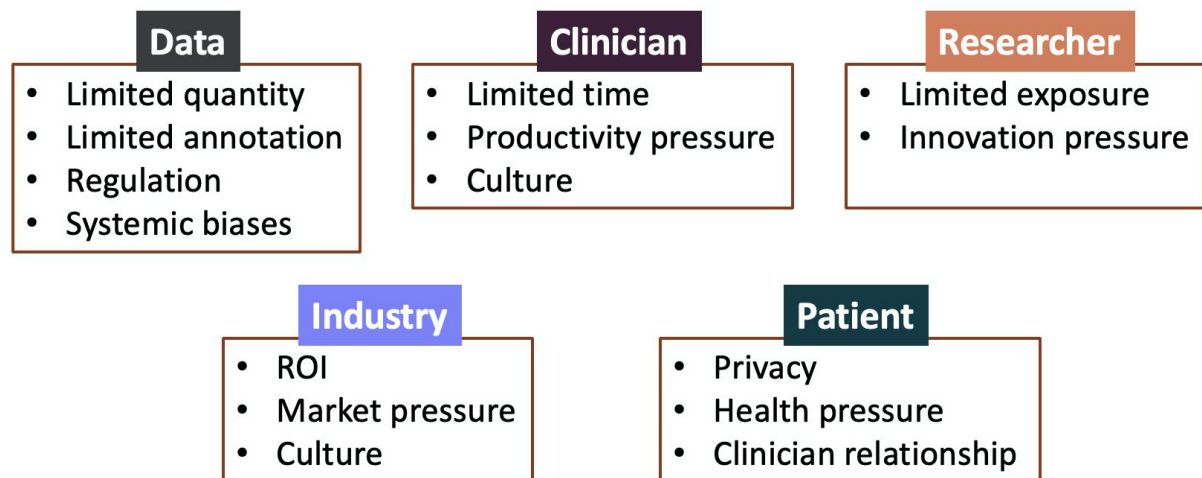


Figure 3. Obstacles to the development and translation of AI in surgery based on key elements and stakeholders such as data, clinicians, researchers (e.g., computer and data scientists), industry, and patients.

If complications of a procedure were to occur as the result of a black box algorithm, it would be increasingly difficult to describe the reason for these complications to the patient. Such concerns have driven an interest in explainable AI - methods that seek to impart at least some sense of meaning to the outputs generated by complex computational processes^[26]. However, incorporation of explainability alone will not solve all problems associated with the use of black box AI approaches, and rigorous assessment of algorithms remains key to responsible and ethical use^[27].

Stakeholder obstacles

Beyond technical obstacles presented by data and methods, behavioral and regulatory incentives play a large role in both positively and negatively influencing the progress of AI for HPB surgery. Clinicians, particularly with growing workforce shortages in healthcare, are subject to increasing demands on clinical productivity. Such pressure can reduce the time and bandwidth to engage in important aspects of professional development and self-improvement, including academic contributions and continuing medical education. Furthermore, while establishing a culture of safety has become the focus of most surgical departments, fear of litigation and unclear hospital policies continue to drive hesitance in systematically engaging in intraoperative video review^[28]. For computer and data scientists, lack of access to the clinical environment may limit their ability to translate their research into clinically meaningful applications. In addition, the academic pressure to produce innovative methods or state-of-the-art results rather than practical applications may limit the translation of some of these methods within an academic environment. Thus, collaboration with industry partners may be helpful in getting these technologies from the laboratory to the clinic and the operating room. However, the need for a significant return on investment as well as business practices that may limit collaboration with competing companies could limit the availability and accessibility of these technologies. Finally and most importantly, patients are the key drivers of AI technologies as patient data is used to develop AI models and the intended beneficiaries of AI models are patients for whom we hope to improve care. It is, therefore, important to ensure that patient privacy and autonomy are respected and that patient perspectives are incorporated into the delivery models for AI.

HPB SURGEONS' ROLE IN ADVANCING AI

While much of the work in developing AI applications in HPB surgery may seem technical, surgeons play a key role in ensuring that these technologies are translated from the research environment into clinical

practice. Perhaps the most straightforward role that surgeons can play is in the collection and annotation of intraoperative data. As previously discussed, AI algorithms often require large amounts of data to optimize their performance. For computer vision applications, this data comes primarily in the form of operative video. Surgeons should seek to clarify institutional rules and policies surrounding the recording, storage, and use of surgical video. Increasing the pool of available data for training AI algorithms could help to improve the generalizability of these algorithms and address some elements of representativeness bias in existing public datasets.

In addition to collecting operative data, surgeons of all training levels are needed to help label or annotate the data. As previously discussed, supervised and semi-supervised learning approaches are dependent on having at least some (if not copious amounts of) labeled data from which algorithms can learn. A growing literature of annotation guides and surgical process ontologies has made the annotation process more accessible and somewhat more standardized^[22,29]. Furthermore, surgical societies such as the Society of American Gastrointestinal and Endoscopic Surgeons, the Japanese Society for Endoscopic Surgery, and the European Association of Endoscopic Surgery have been engaging in work that trains their members to annotate videos for various types of AI projects.

Perhaps the most important contribution that HPB surgeons can make to the field of AI is to participate in multidisciplinary teams and to educate themselves further about the realities of AI. Computer and data scientists can work wonders to draw insights from data; however, it is important to have clinicians on the team who can place such insights into context. The translation of advances in mathematics, statistics, and computer science to clinical medicine can be difficult, as not all clinical problems fit neatly into a given mathematical approach or set of tools. Clinical expertise can help to identify missing data, inappropriately labeled data, or misattributed predictions. It is, thus, important for clinicians to invest time into understanding key methodological considerations in AI and modeling research to ensure that the literature is rigorously evaluated and interpreted before being applied to patients^[30].

CONCLUSION

Artificial intelligence continues to grow at a rapid pace and its applications to surgery are becoming increasingly appreciated. However, obstacles remain in the further development of clinically applicable, reliable, and verifiable algorithms that can translate to patient care. While we have provided a brief and broad overview of some of the terms, techniques, and applications of AI for HPB surgery, it is important for HPB surgeons to dive more deeply into these topics, critically appraise the literature on AI applications, and partner with computer and data scientists to further advance the field.

DECLARATIONS

Authors' contributions

Study conception, drafting, and revision: Boutros C, Singh V, Hashimoto DA

Study conception, review, and revision: Ocun L, Marks J

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

Daniel Hashimoto is a consultant for Johnson and Johnson Institute. He serves on the board of directors of the Global Surgical AI Collaborative, an independent, non-profit organization that oversees and manages a global data-sharing and analytics platform for surgical data. Jeffrey Marks is a consultant for US Endoscopy and Boston Scientific. Christina Boutros, Vivek Singh, and Lee Ocuin have no relevant conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2022.

REFERENCES

1. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg* 2018;268:70-6. DOI PubMed PMC
2. Maier-Hein L, Eisenmann M, Sarikaya D, et al. Surgical data science - from concepts toward clinical translation. *Med Image Anal* 2022;76:102306. DOI PubMed PMC
3. Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial intelligence in anesthesiology: current techniques, clinical applications, and limitations. *Anesthesiology* 2020;132:379-94. DOI PubMed PMC
4. Hashimoto DA, Ward TM, Meireles OR. The role of artificial intelligence in surgery. *Adv Surg* 2020;54:89-101. DOI PubMed
5. Alapatt D, Mascagni P, Vardazaryan A, et al. Temporally constrained neural networks (TCNN): a framework for semi-supervised video semantic segmentation. *Comput Vis Pattern Recognit* 2021. DOI
6. Bektaş M, Zonderhuis BM, Marquering HA, Costa Pereira J, Burchell GL, van der Peet DL. Artificial intelligence in hepatopancreaticobiliary surgery: a systematic review. *Art Int Surg* 2022;2:132-43. DOI
7. Stam WT, Goedknecht LK, Ingwersen EW, Schoonmade LJ, Bruns ERJ, Daams F. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: a systematic review. *Surgery* 2022;171:1014-21. DOI PubMed
8. Merath K, Hyer JM, Mehta R, et al. Use of machine learning for prediction of patient risk of postoperative complications after liver, pancreatic, and colorectal surgery. *J Gastrointest Surg* 2020; 24:1843-51. DOI PubMed
9. Borakati A, Banu Z, Raptis D, et al. New onset diabetes after partial pancreatectomy: development of a novel predictive model using machine learning. *HPB* 2021;23:S814. DOI
10. Al-Haddad MA, Friedlin J, Kesterson J, et al. Natural language processing for the development of a clinical registry: a validation study in intraductal papillary mucinous neoplasms. *HPB* 2010;12:688-95. DOI
11. Roch AM, Mehrabi S, Krishnan A, et al. Automated pancreatic cyst screening using natural language processing: a new tool in the early detection of pancreatic cancer. *HPB (Oxford)* 2015;17:447-53. DOI PubMed PMC
12. Tunstall L, von Werra L, Wolf T. Natural language processing with transformers. Available from: <https://www.amazon.com/Natural-Language-Processing-Transformers-Revised/dp/1098136799> [Last accessed on 29 Dec 2022].
13. Payne TH, Alonso WD, Markiel JA, et al. Using voice to create inpatient progress notes: effects on note timeliness, quality, and physician satisfaction. *JAMIA Open* 2018;1:218-26. DOI PubMed PMC
14. Ward TM, Mascagni P, Ban Y, et al. Computer vision in surgery. *Surgery* 2021;169:1253-1256. DOI PubMed
15. Kenngott HG, Wagner M, Gondan M, et al. Real-time image guidance in laparoscopic liver surgery: first clinical experience with a guidance system based on intraoperative CT imaging. *Surg Endosc* 2014;28:933-40. DOI PubMed
16. Haouchine N, Cotin S, Peterlik I, et al. Impact of soft tissue heterogeneity on augmented reality for liver surgery. *IEEE Trans Vis Comput Graph* 2015;21:584-97. DOI PubMed
17. Giannone F, Felli E, Cherkaoui Z, Mascagni P, Pessaux P. Augmented reality and image-guided robotic liver surgery. *Cancers* 2021;13. DOI PubMed PMC
18. Mascagni P, Alapatt D, Urade T, et al. A computer vision platform to automatically locate critical events in surgical videos: documenting safety in laparoscopic cholecystectomy. *Ann Surg* 2021;274:e93-5. DOI PubMed
19. 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* 2022; 275:955-961. DOI PubMed
20. 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* 2022;276:363-9. DOI PubMed PMC
21. Ward TM, Hashimoto DA, Ban Y, Rosman G, Meireles OR. Artificial intelligence prediction of cholecystectomy operative course

- from automated identification of gallbladder inflammation. *Surg Endosc* 2022;36:6832-40. [DOI](#) [PubMed](#)
22. Meireles OR, Rosman G, Altieri MS, et al; SAGES Video Annotation for AI Working Groups. SAGES consensus recommendations on an annotation framework for surgical video. *Surg Endosc* 2021;35:4918-29. [DOI](#) [PubMed](#)
 23. Ward TM, Fer DM, Ban Y, Rosman G, Meireles OR, Hashimoto DA. Challenges in surgical video annotation. *Comput Assist Surg (Abingdon)* 2021;26:58-68. [DOI](#) [PubMed](#)
 24. Athey S. Beyond prediction: using big data for policy problems. *Science* 2017;355:483-5. [DOI](#) [PubMed](#)
 25. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. *NPJ Digit Med* 2018;1:5. [DOI](#) [PubMed](#) [PMC](#)
 26. Gordon L, Grantcharov T, Rudzicz F. Explainable artificial intelligence for safe intraoperative decision support. *JAMA Surg* 2019;154:1064-5. [DOI](#) [PubMed](#)
 27. Ghassemi M, Oakden-rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digital Health* 2021;3:e745-50. [DOI](#) [PubMed](#)
 28. Mazer L, Varban O, Montgomery JR, Awad MM, Schulman A. Video is better: why aren't we using it? *Surg Endosc* 2022;36:1090-7. [DOI](#) [PubMed](#)
 29. Gibaud B, Forestier G, Feldmann C, et al. Toward a standard ontology of surgical process models. *Int J Comput Assist Radiol Surg* 2018;13:1397-408. [DOI](#) [PubMed](#)
 30. Smeden M. A very short list of common pitfalls in research design, data analysis, and reporting. *PRiMER* 2022;6:26. [DOI](#) [PubMed](#) [PMC](#)