

Artificial Intelligence in Surgery: Promises and Perils

Daniel A. Hashimoto, MD, MS,* Guy Rosman, PhD,† Daniela Rus, PhD,† and Ozanan R. Meireles, MD, FACS*

Objective: The aim of this review was to summarize major topics in artificial intelligence (AI), including their applications and limitations in surgery. This paper reviews the key capabilities of AI to help surgeons understand and critically evaluate new AI applications and to contribute to new developments.

Summary Background Data: AI is composed of various subfields that each provide potential solutions to clinical problems. Each of the core subfields of AI reviewed in this piece has also been used in other industries such as the autonomous car, social networks, and deep learning computers.

Methods: A review of AI papers across computer science, statistics, and medical sources was conducted to identify key concepts and techniques within AI that are driving innovation across industries, including surgery. Limitations and challenges of working with AI were also reviewed.

Results: Four main subfields of AI were defined: (1) machine learning, (2) artificial neural networks, (3) natural language processing, and (4) computer vision. Their current and future applications to surgical practice were introduced, including big data analytics and clinical decision support systems. The implications of AI for surgeons and the role of surgeons in advancing the technology to optimize clinical effectiveness were discussed.

Conclusions: Surgeons are well positioned to help integrate AI into modern practice. Surgeons should partner with data scientists to capture data across phases of care and to provide clinical context, for AI has the potential to revolutionize the way surgery is taught and practiced with the promise of a future optimized for the highest quality patient care.

Keywords: artificial intelligence, clinical decision support, computer-assisted surgery, computer vision, deep learning, machine learning, natural language processing, neural networks, surgery

(*Ann Surg* 2018;xx:xxx–xxx)

Artificial intelligence (AI) can be loosely defined as the study of algorithms that give machines the ability to reason and perform cognitive functions such as problem solving, object and word recognition, and decision-making.¹ Previously thought to be science fiction, AI has increasingly become the topic of both popular and academic literature as years of research have finally built to thresholds of knowledge that have rapidly generated practical applications, such as International Business Machine's (Armonk, NY) and Watson and Tesla's (Palo Alto, CA) autopilot.²

From the *Department of Surgery, Massachusetts General Hospital, Boston, MA; and †Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Boston, MA.

Disclosures: Daniel Hashimoto is financially supported by NIH grant T32DK007754-16A1 and the Massachusetts General Hospital Edward D. Churchill Fellowship. Drs. Ozanan Meireles and Daniel Hashimoto have grant funding from the Natural Orifice Surgery Consortium for Assessment and Research (NOSCAR) for research related to computer vision in endoscopic surgery. Guy Rosman and Daniela Rus are grant funded by Toyota Research Institute for research on autonomous vehicles. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, NOSCAR, or TRI.

The authors report no conflict of interests.

Reprints: Daniel A. Hashimoto, MD, MS, Department of Surgery, Massachusetts General Hospital, 55 Fruit Street, GRB 425, Boston, MA 02114.

E-mail: dahashimoto@mgh.harvard.edu.

Copyright © 2018 Wolters Kluwer Health, Inc. All rights reserved.

ISSN: 0003-4932/16/XXXX-0001

DOI: 10.1097/SLA.0000000000002693

Stories of man-versus-machine, such as that of John Henry working to death to outperform the steam-powered hammer,³ demonstrate how machines have long been feared yet ultimately both accepted and eagerly anticipated. Society proceeded to integrate simple machines into human workflow, and the resulting Industrial Revolution yielded a massive shift in productivity and quality of life. Similarly, AI has inspired awe and struck fear in people who now face a technology that can not only outperform but also potentially outthink its creators.

With the Information Age, a shift in workflow and productivity similar to that of the Industrial Revolution has begun; and surgery stands to gain from the current explosion of information technology. However, as with many emerging technologies, the true promise of AI can be lost in its hype.⁴

It is, therefore, important for surgeons to have a foundation of knowledge of AI to understand how it may impact health care and to consider ways in which they may interact with this technology. This review provides an introduction to AI by highlighting 4 core subfields—(1) machine learning, (2) natural language processing, (3) artificial neural networks, and (4) computer vision—their limitations, and future implications for surgeons.

SUBFIELDS IN AI

AI's roots are found across multiple fields, including robotics, philosophy, psychology, linguistics, and statistics.⁵ Major advances in computer science, such as improvements in processing speed and power, have functioned as a catalyst to allow for the base technologies required for the advent of AI. The growing popularity of AI across many different industries has attracted venture capital investment up to \$5 billion in 2016 alone.⁶ Much of the current attention on AI has focused on the 4 core subfields introduced below.

Machine Learning

Machine learning (ML) enables machines to learn and make predictions by recognizing patterns. Traditional computer programs are explicitly programmed with a desired behavior (eg, when the user clicks an icon, a new program opens). ML allows a computer to utilize partial labeling of the data (supervised learning) or the structure detected in the data itself (unsupervised learning) to explain or make predictions about the data without explicit programming (Fig. 1). Supervised learning is useful for training an ML algorithm to predict a known result or outcome, whereas unsupervised learning is useful in searching for patterns within data.⁷

A third category within machine learning is reinforcement learning, where a program attempts to accomplish a task (eg, driving a car, inferring medical decisions) while learning from its own successes and mistakes.⁸ One can conceptualize reinforcement learning as the computer science equivalent of operant conditioning⁹ and is useful for automated tuning of predictions or actions, such as controlling an artificial pancreas system to fine-tune the measurement and delivery of insulin to diabetic patients.¹⁰

ML is particularly useful for identifying subtle patterns in large datasets—patterns that may be imperceptible to humans performing manual analyses—by using techniques that allow for more indirect and complex nonlinear relationships and multivariate effects

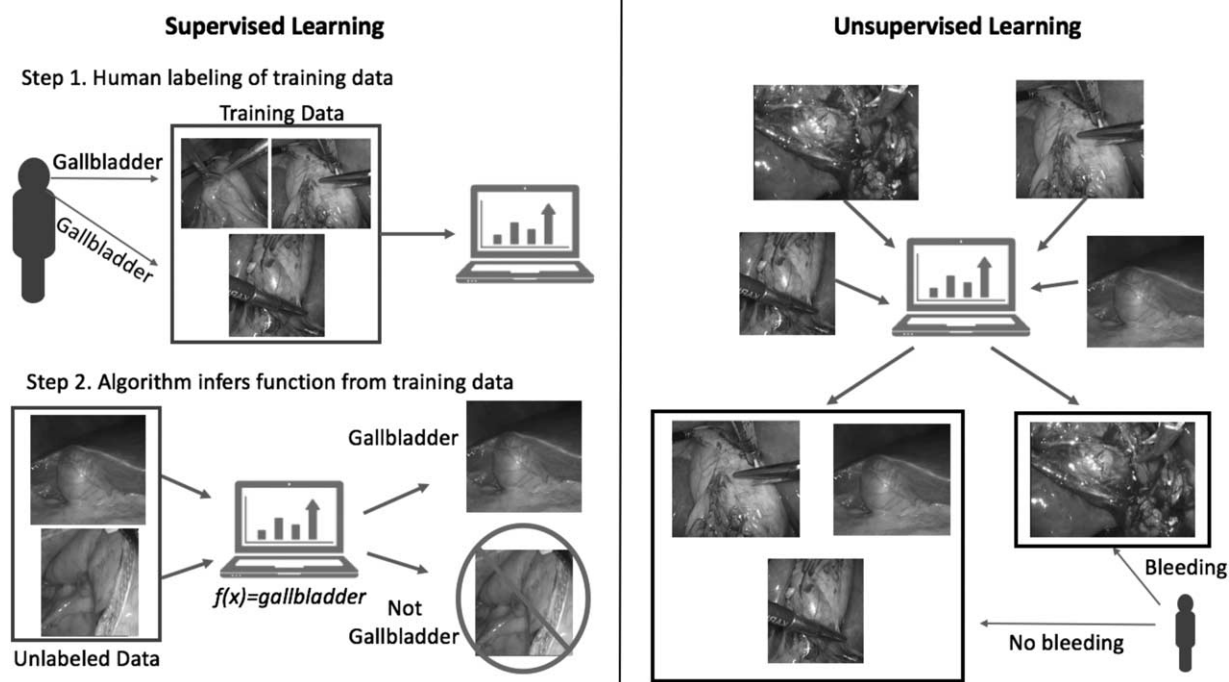


FIGURE 1. In supervised learning, human labeled data are fed to a machine learning algorithm to teach the computer a function, such as recognizing a gallbladder in an image or detecting a complication in a large claims database. In unsupervised learning, unlabeled data are fed to a machine learning algorithm, which then attempts to find a hidden structure to the data, such as identifying bright red (eg, bleeding) as different from nonbleeding tissue.

than conventional statistical analysis.^{11,12} ML has outperformed logistic regression for prediction of surgical site infections (SSIs) by building nonlinear models that incorporate multiple data sources, including diagnoses, treatments, and laboratory values.¹³

Furthermore, multiple algorithms working together (ensemble ML) can be used to calculate predictions at accuracy levels thought to be unattainable with conventional statistics.¹⁴ For example, by analyzing patterns of diagnostic and therapeutic data (including surgical resection) in the Surveillance, Epidemiology and End Results (SEER) cancer registry and comparing data to Medicare claims, ensemble ML with random forests, neural networks, and lasso regression was able to predict patient lung cancer staging by using International Classification of Diseases (ICD)-9 claims data alone with 93% sensitivity, 92% specificity, and 93% accuracy, outperforming a decision tree approach based on clinical guidelines alone (53% sensitivity, 89% specificity, and 72% accuracy).¹⁵

Natural Language Processing

Natural language processing (NLP) is a subfield that emphasizes building a computer's ability to understand human language and is crucial for large-scale analyses of content such as electronic medical record (EMR) data, especially physicians' narrative documentation. To achieve human-level understanding of language, successful NLP systems must expand beyond simple word recognition to incorporate semantics and syntax into their analyses.¹⁶

Rather than relying on codified classifications such as ICD codes, NLP enables machines to infer meaning and sentiment from unstructured data (eg, prose written in the history of present illness or in a physician's assessment and plan). NLP allows clinicians to write more naturally rather than having to input specific text sequences or

select from menus to allow a computer to recognize the data. NLP has been utilized for large-scale database analysis of the EMR to detect adverse events and postoperative complications from physician documentation,^{17,18} and many EMR systems now incorporate NLP—for example, to achieve automated claims coding—into their underlying software architecture to improve workflow or billing.¹⁹

In surgical patients, NLP has been used to automatically comb through EMRs to identify words and phrases in operative reports and progress notes that predicted anastomotic leak after colorectal resections. Many of its predictions reflected simple clinical knowledge that a surgeon would have (eg, operation type and difficulty), but the algorithm was also able to adjust predictive weights of phrases describing patients (eg, irritated, tired) relative to the postoperative day to achieve predictions of leak with a sensitivity of 100% and specificity of 72%.²⁰ The ability of algorithms to self-correct can increase the utility of their predictions as datasets grow to become more representative of a patient population.

Artificial Neural Networks

Artificial neural networks, a subfield of ML, are inspired by biological nervous systems and have become of paramount importance in many AI applications. Neural networks process signals in layers of simple computational units (neurons); connections between neurons are then parameterized via weights that change as the network learns different input–output maps corresponding to tasks such as pattern/image recognition and data classification (Fig. 2).⁷

Deep learning networks are neural networks composed of many layers and are able to learn more complex and subtle patterns than simple 1- or 2-layer neural networks.²¹

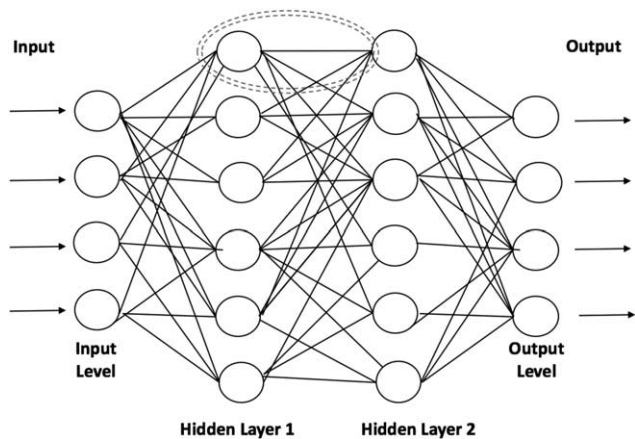


FIGURE 2. Artificial neural networks are composed of many computational units known as “neurons” (dotted red circle) that receive data inputs (similar to dendrites in biological neurons), perform calculations, and transmit output (similar to axons) to the next neuron. Input level neurons receive data, whereas hidden layer neurons (many different hidden layers can be used) conduct the calculations necessary to analyze the complex relationships in the data. Hidden layer neurons then send the data to an output layer that provides the final version of the analysis for interpretation.

Clinically, ANNs have significantly outperformed more traditional risk prediction approaches. For example, an ANN’s sensitivity (89%) and specificity (96%) outperformed APACHE II sensitivity (80%) and specificity (85%) for prediction of pancreatitis 76 hours after admission.²² By using clinical variables such as patient history, medications, blood pressure, and length of stay, ANNs, in combination with other ML approaches, have yielded predictions of in-hospital mortality after open abdominal aortic aneurysm repair with sensitivity of 87%, specificity of 96.1%, and accuracy of 95.4%.²³

Computer Vision

Computer vision describes machine understanding of images and videos, and significant advances have resulted in machines achieving human-level capabilities in areas such as object and scene recognition.²⁴ Important healthcare-related work in computer vision includes image acquisition and interpretation in axial imaging with applications, including computer-aided diagnosis, image-guided surgery, and virtual colonoscopy.²⁵ Initially influenced by statistical signal processing, the field has recently shifted significantly toward

more data-intensive ML approaches, such as neural networks,²⁶ with adaptation into new applications.

Utilizing ML approaches, current work in computer vision is focusing on higher level concepts such as image-based analysis of patient cohorts, longitudinal studies, and inference of more subtle conditions such as decision-making in surgery. For example, real-time analysis of laparoscopic video has yielded 92.8% accuracy in automated identification of the steps of a sleeve gastrectomy and noted missing or unexpected steps.²⁷ With 1-minute of high-definition surgical video estimated to contain 25 times the amount of data found in a high-resolution computed tomography image,²⁸ video could contain a wealth of actionable data.^{29,30} Thus, although predictive video analysis is in its infancy, such work provides proof-of-concept that AI can be leveraged to process massive amounts of surgical data to identify or predict adverse events in real-time for intraoperative clinical decision support (Fig. 3).

SYNERGY ACROSS AI AND OTHER FIELDS

The promise of AI lies in applications that combine aspects of each of the above subfields with other elements of computing such as database management and signal processing.⁷ The increasing potential of AI in surgery is analogous to other recent technological developments (eg, mobile phones, cloud computing) that have arisen from the intersection of hyper-cycle advances in both hardware and software (ie, as hardware advances, so too does software and vice versa).

Synergy between fields is also important in expanding the applications of AI. Combining NLP and computer vision, Google (Mountain View, CA) Image Search is able to display relevant pictures in response to a textual query such as a word or phrase. Furthermore, neural networks, specifically deep learning, now form a significant part of the architecture underlying various AI systems. For example, deep learning in NLP has allowed for significant improvements in the accuracy of translation (60% more accurate translation by Google Translate³¹), whereas its use in computer vision has resulted in greater accuracy of classification of images (42% more accurate image classification by AlexNet³²).

Clinical applications of such work include the successful utilization of deep learning to create a computer vision algorithm for the classification of smartphone images of benign and malignant skin lesions at an accuracy level equivalent to dermatologists.³³ NLP and ML analyses of postoperative colorectal patients demonstrated that prediction of anastomotic leaks improved to 92% accuracy when different data types were analyzed in concert instead of individually (accuracy of vital signs—65%; laboratory values—74%; text data—83%).³⁴

Early attempts at using AI for technical skills augmentation focused on small feats such as task deconstruction and autonomous performance of simple tasks (eg, suturing, knot-tying).^{35,36} Such

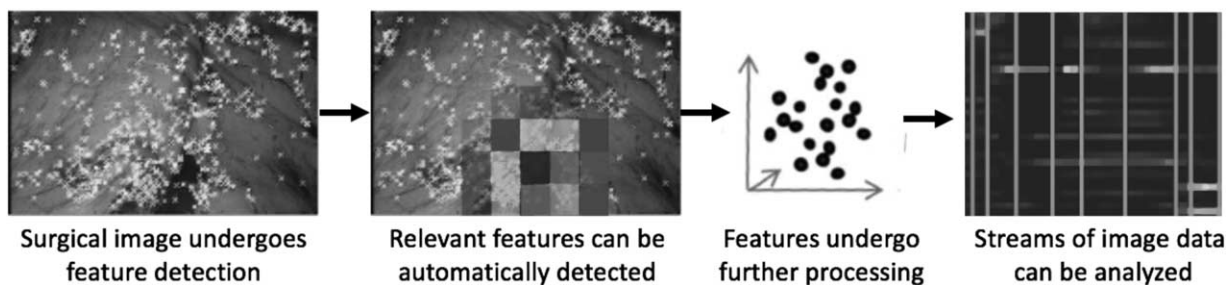


FIGURE 3. Computer vision utilizes mathematical techniques to analyze visual images or video streams as quantifiable features such as color, texture, and position that can then be used within a dataset to identify statistically meaningful events such as bleeding.

efforts have been critical to establishing a foundation of knowledge for more complex AI tasks.³⁷ For example, the Smart Tissue Autonomous Robot (STAR) developed by Johns Hopkins University was equipped with algorithms that allowed it to match or outperform human surgeons in autonomous ex vivo and in vivo bowel anastomosis in animal models.³⁸

Although truly autonomous robotic surgery will remain out of reach for some time, synergy across fields will likely accelerate the capabilities of AI in augmenting surgical care. For AI, much of its clinical potential is in its ability to analyze combinations of structured and unstructured data (eg, EMR notes, vitals, laboratory values, video, and other aspects of “big data”) to generate clinical decision support. Each type of data could be analyzed independently or in concert with different types of algorithms to yield innovations.

The true potential of AI remains to be seen and could be difficult to predict at this time. Synergistic reactions between different technologies can lead to unanticipated revolutionary technology; for example, recent synergistic combinations of advanced robotics, computer vision, and neural networks led to the advent of autonomous cars. Similarly, independent components within AI and other fields could combine to create a force multiplier effect with unanticipated changes to healthcare delivery. Therefore, surgeons should be engaged in assessing the quality and applicability of AI advances to ensure appropriate translation to the clinical sector.

Limitations of AI

As with any new technology, AI and each of its subfields are susceptible to unrealistic expectations from media hype that can lead to significant disappointment and disillusionment.³⁹ AI is not a “magic bullet” that can yield answers to all questions. There are instances where traditional analytical methods can outperform ML⁴⁰ or where the addition of ML does not improve on its results.⁴¹ As with any scientific endeavor, use of AI hinges on whether the correct scientific question is being asked and whether one has the appropriate data to answer that question.

ML provides a powerful tool with which to uncover subtle patterns in data. It excels at detecting patterns and demonstrating correlations that may be missed by traditional methods, and these results can then be used by investigators to uncover new clinical questions or generate novel hypotheses about surgical diseases and management.^{42,43} However, there are both costs and risks to utilizing ML incorrectly.

The outputs of ML and other AI analyses are limited by the types and accuracy of available data. Systematic biases in clinical data collection can affect the type of patterns AI recognizes or the predictions it may make,^{44,45} and this can especially affect women and racial minorities due to long-standing underrepresentation in clinical trial and patient registry populations.^{46–48} Supervised learning is dependent on labeling of data (such as identification of variables currently used in surgery-specific patient registries) which can be expensive to gather, and poorly labeled data will yield poor results. A publically available National Institutes of Health (NIH) dataset of chest x-rays and reports has been utilized to generate AI capable of generating diagnoses of chest x-rays. NLP was used to mine radiology reports to generate labels for chest x-rays, and these labels were used to train a deep learning network to recognize pathology on images with particularly good accuracy in identifying a pneumothorax.⁴⁹ However, an in-depth analysis of the dataset by Oakden-Rayner⁵⁰ revealed that some of the results may have been from improperly labeled data. Most of the x-rays labeled as pneumothorax also had a chest tube present, raising concern that the network was identifying chest tubes rather than pneumothoraces as intended.

An important concern regarding AI algorithms involves their interpretability,⁵¹ for techniques such as neural networks are based on a “black box” design.⁵² Although the automated nature of neural networks allows for detection of patterns missed by humans, human scientists are left with little ability to assess how or why such patterns were discerned by the computer. Medicine has been quick to recognize that the accountability of algorithms, the safety/verifiability of automated analyses, and the implications of these analyses on human–machine interactions can impact the utility of AI in clinical practice.⁵³ Such concerns have hindered the use of AI algorithms in many applicative fields from medicine to autonomous driving and have pushed data scientists to improve the interpretability of AI analyses.^{54,55} However, many of these efforts remain in their infancy, and surgeon input early in the design of AI algorithms may be helpful in improving accountability and interpretability of big data analyses.

Furthermore, despite advances in causal inference, AI cannot yet determine causal relationships in data at a level necessary for clinical implementation nor can it provide an automated clinical interpretation of its analyses.⁵⁶ Although big data can be rich with variables, it is poor in providing the appropriate clinical context with which to interpret the data. Human physicians, therefore, must critically evaluate the predictions generated by AI and interpret the data in clinically meaningful ways.

Implications for Surgeons

The first widespread uses of AI are likely to be in the form of computer-augmentation of human performance. Clinician–machine interaction has already been demonstrated to augment decision-making. Pathologists have utilized AI to decrease their error rate in recognizing cancer-positive lymph nodes from 3.4% to 0.5%.⁵⁷ Furthermore, by allowing for improved identification of high risk patients, AI can assist surgeons and radiologists in reducing the rate of lumpectomy by 30% in patients whose breast needle biopsies are considered high risk lesions but ultimately found to be benign after surgical excision.⁵⁸

In the future, a surgeon will likely see AI analysis of population and patient-specific data augmenting each phase of care (Fig. 4). Preoperatively, a patient undergoing evaluation for bariatric surgery may be tracking weight, glucose, meals, and activity through mobile applications and fitness trackers, with the data feeding into their EMR.^{59–61} Automated analysis of all preoperative mobile and clinical data could provide a more patient-specific risk score for operative planning and yield valuable predictors for postoperative care. The surgeon could then augment their decision-making intraoperatively based on real-time analysis of intraoperative progress that integrates EMR data with operative video, vital signs, instrument/hand tracking, and electrosurgical energy usage. Intraoperative monitoring of such different types of data could lead to real-time prediction and avoidance of adverse events. Integration of pre-, intra-, and postoperative data could help to monitor recovery and predict complications. After discharge, postoperative data from personal devices could continue to be integrated with data from their hospitalization to maximize weight loss and resolution of obesity-related comorbidities.⁶² Such an example could be applied to any type of surgical care with the potential for truly patient-specific, patient-centered care.

AI could be utilized to augment sharing of knowledge through the collection of massive amounts of operative video and EMR data across many surgeons around the world to generate a database of practices and techniques that can be assessed against outcomes. Video databases could use computer vision to capture rare cases or anatomy, aggregating and integrating data across pre-, intra-, and postoperative phases of care.^{63,64} Such powerful analyses could

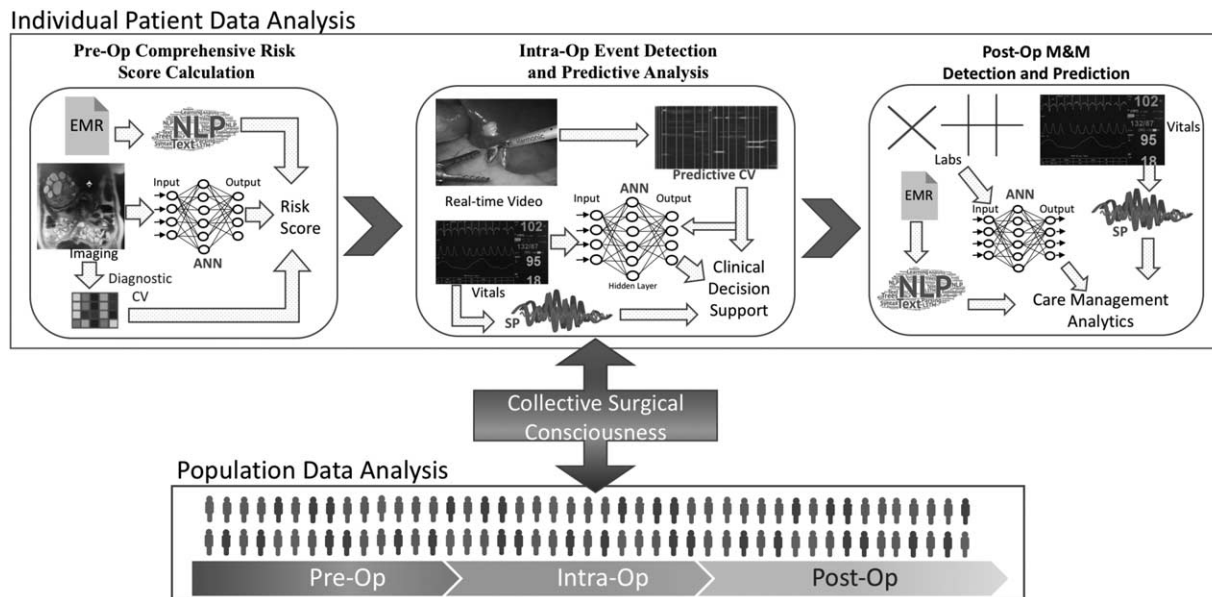


FIGURE 4. Integration of multimodal data with AI can augment surgical decision-making across all phases of care both at the individual patient and at the population level. An integrated AI serving as a “collective surgical consciousness” serves as the conduit to add individual patient data to a population dataset while drawing from population data to provide clinical decision support during individual cases. ANN indicates artificial neural network; CV, computer vision; NLP, natural language processing; SP, signal processing.

create truly disruptive innovation in generating and validating evidence-based best practices to improve care quality.

The Surgeon’s Role

With big data analytics predicted to yield annual healthcare savings between \$300 and \$450 billion annually in the United States alone,⁶⁵ there is great economic incentive to incorporate AI and big data into multiple elements of our healthcare system. Surgeons are uniquely positioned to help drive these innovations rather than passively waiting for the technology to become useful.

As lack of data can limit the predictions made by AI, surgeons should seek to expand involvement in clinical data registries to ensure all patients are included. These can include registries at the local, national, or international levels. As data cleaning techniques improve, registries could become linked to expand their utility and increase the availability of clinical, genomic, proteomic, radiographic, and pathologic data.

Surgeons, as the key stakeholders in adoption of AI-based technologies for surgical care, should seek opportunities to partner with data scientists to capture novel forms of clinical data and help generate meaningful interpretations of that data.⁶⁶ Surgeons have the clinical insight that can guide data scientists and engineers to answer the right questions with the right data, whereas engineers can provide automated, computational solutions to data analytics problems that would otherwise be too costly or time-consuming for manual methods.

Technology-based dissemination of surgical practice can empower every surgeon with the ability to improve the quality of global surgical care. Given that research has demonstrated that surgical technique and skill correlates to patient outcomes,^{67,68} AI could help pool surgical experience—similar to efforts in genomics and biobanks⁶⁹—to bring the decision-making capabilities and techniques of the global surgical community into every operation.

Big data could be leveraged to create a “collective surgical consciousness” that carries the entirety of the field’s knowledge, leading to technology-augmented real-time clinical decision support, such as intraoperative, GPS-like guidance.

Surgeons can provide value to data scientists by imparting their understanding of the relevance and importance of the relationship between seemingly simple topics, such as anatomy and physiology, to more complex phenomena, such as a disease pathophysiology, operative course, or postoperative complications. These types of relationships are important to appropriately model and predict clinical events, and they are critical to improving the interpretability of ML approaches. Surgeons and engineers alike should demand transparency and interpretability in algorithms so that AI can be held accountable for its predictions and recommendations. With patients’ lives at stake, the surgical community should expect automated systems that augment human capabilities to provide care to at least meet, if not exceed, the standards to which clinicians and scientists are held.

Surgeons are ultimately the ones providing clinical information to patients and will have to establish a patient communication framework through which to relay the data made accessible by AI.⁷⁰ An understanding of AI will be key to appropriately conveying the results of complex analyses such as risk predictions, prognostications, and treatment algorithms to patients within the appropriate clinical context.^{4,71}

Working with patients, surgeons should develop and deliver the narrative behind optimal utilization of AI in patient care, avoiding complications that can arise when external forces (eg, regulators, administrators) mandate implementation of new technologies⁷² without fully evaluating potential impacts on those who would use the technology most. If appropriately developed and implemented, AI has the potential to revolutionize the way surgery is taught and practiced with the promise of a future optimized for the highest quality patient care.

CONCLUSIONS

AI is expanding its footprint in clinical systems ranging from databases to intraoperative video analysis. The unique nature of surgical practice leaves surgeons well-positioned to help usher in the next phase of AI, one focused on generating evidence-based, real-time clinical decision support designed to optimize patient care and surgeon workflow.

REFERENCES

- Bellman R. *An Introduction to Artificial Intelligence: Can Computers Think?* San Francisco: Boyd & Fraser Pub Co; 1978.
- Lewis-Kraus G. *The great AI awakening.* New York Times; 2016.
- Keats EJ. *John Henry.* New York: Pantheon: An American Legend; 1965.
- Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N Engl J Med.* 2017;376:2507–2509.
- Buchanan BG. A (very) brief history of artificial intelligence. *AI Mag.* 2005;26:53.
- CBInsights. The 2016 AI Recap: Startups See Record High in Deals and Funding 2017. Available at: <https://www.cbinsights.com/blog/artificial-intelligence-startup-funding/>. Accessed June 4, 2017.
- Deo RC. Machine learning in medicine. *Circulation.* 2015;132:1920–1930.
- Sutton RS, Barto AG. *Reinforcement Learning: An Introduction.* Vol. 1. Cambridge: MIT Press; 1998.
- Skinner BF. *The behaviour of organisms: An experimental analysis.* Oxford: Appleton-Century; 1938.
- Bothe MK, Dickens L, Reichel K, et al. The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. *Expert Rev Med Devices.* 2013;10:661–673.
- Miller RA, Pople HEJ, Myers JD, et al. An experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med.* 1982;307:468–476.
- Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform.* 2006;2:59.
- Soguero-Ruiz C, Fei WM, Jenssen R, et al. Data-driven temporal prediction of surgical site infection. *AMIA Annu Symp Proc.* 2015;2015:1164–1173.
- Wang PS, Walker A, Tsuang M, et al. Strategies for improving comorbidity measures based on Medicare and Medicaid claims data. *J Clin Epidemiol.* 2000;53:571–578.
- Bergquist S, Brooks G, Keating N, et al. Classifying lung cancer severity with ensemble machine learning in health care claims data. *Proc Mach Learn Res.* 2017;68:25–38.
- Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc.* 2011;18:544–551.
- Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA.* 2011;306:848–855.
- Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc.* 2005;12:448–457.
- Friedman C, Shagina L, Lussier Y, et al. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc.* 2004;11:392–402.
- Soguero-Ruiz C, Hindberg K, Rojo-Alvarez JL, et al. Support vector feature selection for early detection of anastomosis leakage from bag-of-words in electronic health records. *IEEE J Biomed Health Inform.* 2016;20:1404–1415.
- Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput.* 2006;18:1527–1554.
- Modifi R, Duff MD, Madhavan KK, et al. Identification of severe acute pancreatitis using an artificial neural network. *Surgery.* 2007;141:59–66.
- Monsalve-Torra A, Ruiz-Fernandez D, Marin-Alonso O, et al. Using machine learning methods for predicting in-hospital mortality in patients undergoing open repair of abdominal aortic aneurysm. *J Biomed Inform.* 2016;62:195–201.
- Szeliski R. *Computer Vision: Algorithms and Applications.* London: Springer-Verlag London; 2010.
- Kenngott HG, Wagner M, Nickel F, et al. Computer-assisted abdominal surgery: new technologies. *Langenbecks Arch Surg.* 2015;400:273–281.
- Egmont-Petersen M, de Ridder D, Handels H. Image processing with neural networks—a review. *Pattern Recognit.* 2002;35:2279–2301.
- Volkov M, Hashimoto DA, Rosman G, et al. Machine Learning and Coresets for Automated Real-Time Video Segmentation of Laparoscopic and Robot-Assisted Surgery. In: IEEE International Conference on Robotics and Automation, Singapore; 2017:754–759.
- Natarajan P, Frenzel JC, Smaltz DH. *Demystifying Big Data and Machine Learning for Healthcare.* Boca Raton: CRC Press; 2017.
- Bonrath EM, Gordon LE, Grantcharov TP. Characterising “near miss” events in complex laparoscopic surgery through video analysis. *BMJ Qual Saf.* 2015;24:516–521.
- Grenda TR, Pradarelli JC, Dimick JB. Using surgical video to improve technique and skill. *Ann Surg.* 2016;264:32–33.
- Johnson M, Schuster M, Le QV, et al. Google’s multilingual neural machine translation system: enabling zero-shot translation. *Transactions of the Association for Computational Linguistics.* 2017;5:339–351.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems.* 2012;25:1097–1105.
- Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542:115–118.
- Soguero-Ruiz C, Hindberg K, Mora-Jimenez I, et al. Predicting colorectal surgical complications using heterogeneous clinical data and kernel methods. *J Biomed Inform.* 2016;61:87–96.
- DiPietro R, Lea C, Malpani A, et al. Recognizing surgical activities with recurrent neural networks. In: Ourselin S, Joskowicz L, Sabuncu M, et al., eds. *Medical Image Computing and Computer-Assisted Intervention.* Cham: Springer International Publishing; 2016. 551–558.
- Zappella L, Béjar B, Hager G, et al. Surgical gesture classification from video and kinematic data. *Med Image Anal.* 2013;17:732–745.
- Moustris GP, Hiriadis SC, Deliparaschos KM, et al. Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. *Int J Med Robot.* 2011;7:375–392.
- Shademan A, Decker RS, Opfermann JD, et al. Supervised autonomous robotic soft tissue surgery. *Sci Transl Med.* 2016;8:337ra64–1337ra.
- Linden A, Fenn J. *Understanding Gartner’s Hype Cycles (Strategic Analysis Report No. R-20-1971).* Stamford: Gartner, Inc; 2003.
- Austin PC, Tu JV, Lee DS. Logistic regression had superior performance compared with regression trees for predicting in-hospital mortality in patients hospitalized with heart failure. *J Clin Epidemiol.* 2010;63:1145–1155.
- Bellman RE. *Adaptive Control Processes: A Guided Tour.* Princeton: Princeton University Press; 2015.
- Rudin C, Dunson D, Irizarry R, et al. Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society. In: American Statistical Association White Paper; 2014.
- Council NR. *Frontiers in Massive Data Analysis.* Washington, DC: National Academies Press; 2013.
- Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ.* 2001;323:42–46.
- Hopewell S, Loudon K, Clarke MJ, et al. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev.* 2009;1:MR000006.
- Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA.* 2004;291:2720–2726.
- Chang AM, Mumma B, Sease KL, et al. Gender bias in cardiovascular testing persists after adjustment for presenting characteristics and cardiac risk. *Acad Emerg Med.* 2007;14:599–605.
- Douglas PS, Ginsburg GS. The evaluation of chest pain in women. *N Engl J Med.* 1996;334:1311–1315.
- Wang X, Peng Y, Lu L, et al. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In: IEEE CVPR, Honolulu, HI; 2017.
- Oakden-Rayner L. *Exploring the ChestXray14 dataset: Problems.* Wordpress: Luke Oakden Rayner; 2017.
- Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: ACM; 2016:1135–1144.
- Sussillo D, Barak O. Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput.* 2013;25:626–649.
- Cabrita F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA.* 2017;318:517–518.

54. Sturm I, Lapuschkin S, Samek W, et al. Interpretable deep neural networks for single-trial EEG classification. *J Neurosci Methods*. 2016;274:141–145.
55. Tan S, Sim KC, Gales M. Improving the interpretability of deep neural networks with stimulated learning. In: Automatic Speech Recognition and Understanding (ASRU), 2015 IEEE Workshop on: IEEE; 2015:617–623.
56. Pearl J. Causality: Models, Reasoning and Inference. 2nd ed. Cambridge, UK: Cambridge University Press; 2009.
57. Wang D, Khosla A, Gargeya R, et al. Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718 2016. Epub ahead of print.
58. Bahl M, Barzilay R, Yedidia AB, et al. High-risk breast lesions: a machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. *Radiology*; 0:170549. Epub ahead of print.
59. Harvey C, Koubek R, Begat V, et al. usability evaluation of a blood glucose monitoring system with a spill-resistant vial, easier strip handling, and connectivity to a mobile app: improvement of patient convenience and satisfaction. *J Diabetes Sci Technol*. 2016;10:1136–1141.
60. Horner GN, Agboola S, Jethwani K, et al. Designing patient-centered text messaging interventions for increasing physical activity among participants with type 2 diabetes: qualitative results from the text to move intervention. *JMIR Mhealth Uhealth*. 2017;5:e54.
61. Ryu B, Kim N, Heo E, et al. Impact of an electronic health record-integrated personal health record on patient participation in health care: development and randomized controlled trial of MyHealthKeeper. *J Med Internet Res*. 2017;19:e401.
62. Elvin-Walsh L, Ferguson M, Collins PF. Nutritional monitoring of patients post-bariatric surgery: implications for smartphone applications. *J Hum Nutr Diet*. 2018;31:141–148.
63. Langerman A, Grantcharov TP. Are we ready for our close-up? Why and how we must embrace video in the OR. *Ann Surg*. 2017;266:934–936.
64. Hashimoto DA, Rosman G, Rus D, et al. Surgical video in the age of big data. *Ann Surg*. 2017 [Epub ahead of print].
65. Groves P, Kayyali B, Knott D, et al. *The 'Big Data' Revolution in Healthcare: Accelerating Value and Innovation*. New York: McKinsey & Co; 2013.
66. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. *JAMA*. 2014;311:2479–2480.
67. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013;369:1434–1442.
68. Scally CP, Varban OA, Carlin AM, et al. Video ratings of surgical skill and late outcomes of bariatric surgery. *JAMA Surg*. 2016;151:e160428.
69. O'Shea P. Future medicine shaped by an interdisciplinary new biology. *Lancet*. 2012;379:1544–1550.
70. Emanuel EJ, Emanuel LL. Four models of the physician-patient relationship. *JAMA*. 1992;267:2221–2226.
71. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375:1216–1219.
72. Miller RH, Sim I. Physicians' use of electronic medical records: barriers and solutions. *Health Aff*. 2004;23:116–126.