Advances in Surgery 54 (2020) 89–101

ADVANCES IN SURGERY

The Role of Artificial Intelligence in Surgery

Daniel A. Hashimoto, MD, MS*, Thomas M. Ward, MD[1], Ozanan R. Meireles, MD[2]

Surgical AI & Innovation Laboratory, Department of Surgery, Massachusetts General Hospital, Harvard Medical School, 15 Parkman Street, WAC460, Boston, MA 02114, USA

Keywords

• Artificial intelligence • Deep learning • Computer vision • Natural language processing • Learning algorithms • Risk prediction

Key points

- Artificial intelligence (AI) as a term has been used indiscriminately by marketers and media, and surgeons should approach usage of the term with healthy skepticism.
- All current forms of AI in surgery are narrow AI, that is, tools designed for use in specific tasks under specific circumstances.
- Three major forms of machine learning are supervised learning, unsupervised learning, and reinforcement learning.

INTRODUCTION

Artificial intelligence (AI) is loosely defined as the study of algorithms that give machines the ability to reason and perform cognitive functions [1]. Although often considered by the wider public to be a field firmly held in the domain of computer science, AI is a field with wide roots ranging from mathematics to statistics to computer science to philosophy to psychology to neurobiology and to linguistics.

The popular conception of AI as represented in film and television is that of a machine capable of mimicking human behavior or intellect–Hal 9000 from *2001: A Space Odyssey*, the Terminator, or J.A.R.V.I.S. and Vision from *The*

[1]Present address: 37 Joy Street, Apartment 7, Boston, MA 02114.

[2]Present address: 500 Atlantic Avenue, Apartment 16C, Boston, MA 02210.

*Corresponding author. *E-mail address:* dahashimoto@mgh.harvard.edu

https://doi.org/10.1016/j.yasu.2020.05.010

Avengers movies. These are fictionalized portrayals of artificial general (also known as strong) intelligence–AI that is capable of tackling any task. In reality, the AI algorithms under development or in use at the moment fall under the category of narrow, or weak, intelligence, that is, AI that is capable of accomplishing very specific tasks.

Highlighting the specific nature of these algorithms, the Food and Drug Administration (FDA) approved the first diagnostic utilization of an AI algorithm in 2018–a program that assists in screening for diabetic retinopathy through automated analysis of images of the fundus [2]. The list of FDA-approved algorithms continues to grow with approved applications in radiology and cardiology and pathology as well. With ongoing development and application of AI technologies in medicine, it is important for clinicians in every field to understand what these technologies are and how they can be leveraged to deliver safer, more efficient, and more cost-effective care. Furthermore, it is important to keep in mind what these technologies are not and to understand the limitations inherent to any tool.

TECHNIQUES IN ARTIFICIAL INTELLIGENCE

The current utilization of the term, AI, has grown unwieldy, particularly when used by marketing departments and the media. Often, AI is used in place of more specific terms, such as machine learning (ML) or deep learning. Thus, to provide an example of a taxonomy of how AI relates to more specific terms, AI can be considered a larger parent field that encompasses subfields like ML, which further encompasses techniques like neural networks and deep learning. Fig. 1 demonstrates a conceptualization of a taxonomy of AI to assist in understanding how some of these topics relate to one another. These subfields and others are quite inter-related, and techniques from one may be subsumed or used in concert with another, depending on the application.

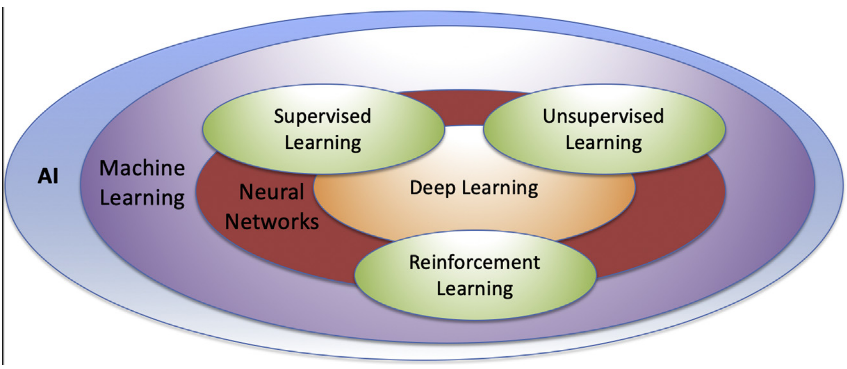

Fig. 1. Relationship of terminology in AI used in this article.

Delving into the details behind each of the various subfields of AI is important for an in-depth understanding of the field; however, for the curious surgeon, an overview of the major techniques utilized in surgical applications of AI should suffice as an introduction to work in the field.

Machine Learning

ML refers to the study of algorithms and statistical models that enable machines to learn and perform tasks. ML algorithms utilize features, or properties within the data, to perform tasks without explicit programming. These tasks traditionally are divided into those that require classification (ie, dividing data into classes) and those that require regression (ie, modeling the relationship between continuous variables). In classic ML, the features are selected or hand-crafted by humans to guide the algorithms in evaluating specific components within the data during its analysis. This is in contrast to neural networks (described later), where features are extracted automatically.

Within ML, the 2 most common learning types are supervised and unsupervised learning. Supervised learning is a task-driven process wherein an algorithm is trained to predict a prespecified output, such as identifying a stop sign or recognizing a cat in a photograph. The "supervised" moniker comes from the need to provide annotated (ie, labeled) data so that it can learn the associations between inputs and the desired output. Thus, data sets are divided into a training set (with labels provided) for learning and a test set (no labels provided) that allows for the assessment of the performance of the algorithm on new data [3,4].

Unlike supervised learning, unsupervised learning does not utilize a prespecified annotation; rather, it draws inferences from unlabeled data to identify patterns and/or structure within a data set. This type of learning can be useful in identifying relationships between groups (eg, clustering) for further hypothesis generation. This can be applied to typical, discrete surgical data, such as patient outcomes databases, or to more unique data sets, such as surgical motion and activity. For example, unsupervised learning has been used to identify high-risk cardiac surgery patients and to automatically identify suturing motion in surgical video [5,6].

A third category of learning is reinforcement learning, a form of unsupervised learning. It is analogous to operant conditioning, where learning occurs through successive attempts via trial and error and rewards/punishments guide the behavior of the model to optimize rewards [3,7].

Perhaps the most famous example of reinforcement learning comes from the Google AlphaGoZero, a reinforcement learning algorithm initially designed to play Go. Unlike prior computer systems designed to play games where the machine was taught a series of moves or was fed past examples of moves played by master players, AlphaGoZero was given only the rules and learned from self-play, becoming one of the top players in the world in 24 hours [8]. Computer mastery of Go, without the input of human knowledge, was previously thought to be a near-impossible task because of the incredibly large number of move combinations possible (googolplex, or $10^{(10^{100})}$).

Despite this impressive feat, expectations must be tempered for the application of such technology in surgery. Although games can be complex, they are defined by well-understood rules that can be navigated efficiently by algorithms. In contrast, medicine often is defined by uncertainty, and data contain much more noise than signal. Often many features are required to appropriately model a medical phenomenon, increasing the dimensionality of a problem and the difficulty of accurately modeling the phenomenon itself.

Because no model describes medical phenomena perfectly, there must be awareness of potential methodological pitfalls, such as overfitting of data. Overfitting describes a model that fits the data on which it was trained too closely, resulting in predictions that are very high but do not generalize well to outside data sets. In other words, the model memorizes the training data set itself instead of modeling the phenomenon. Thus, in addition to testing performance of a model on a test data set split from the original training data, it is preferable to have an independent data set on which to validate model performance and assess its generalizability.

Neural networks

In classic ML, features (ie, variables) are selected (also known as hand-crafted or hand-engineered) by a person to optimize performance at a given task. For example, hand-crafted features in a task to detect a cat could include whiskers and pointy ears. Neural networks, inspired by biological nervous systems, process data in layers of simple computational units that are intended to be analogous to neurons (Fig. 2). Thus, unlike in classic ML, neural networks can

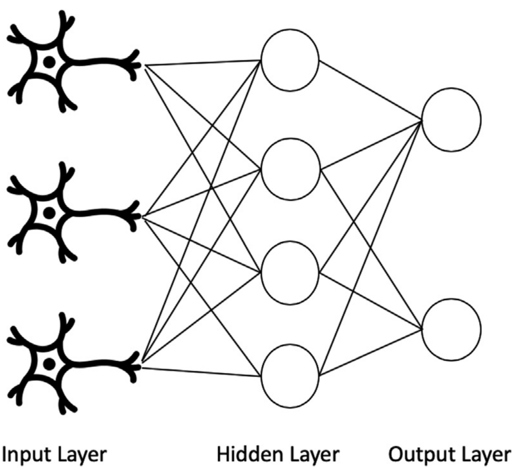

Fig. 2. Illustrative example of a 3-layer neural network consisting of an input layer, a hidden layer, and an output layer.

extract features from data and use them as inputs, adjusting the weights of those features accordingly to be used within an activation function to yield some output [9]. That is, the system automatically, using predetermined mathematical functions, tweaks weights to strengthen/weaken connections within the network to yield the best possible results.

Deep neural networks

Deep neural networks are, effectively, neural networks with more than 3 layers, allowing for learning of more complex patterns than those that are discernible from simple 1-layer or 2-layer networks. As with nondeep neural networks, deep learning selects features that are most likely to yield best results. This technique works particularly well with unstructured data, such as audio, images, and video [10]. Generally, each layer of a deep neural network performs a set of operations to generate a representation of the data that then is fed feeds to the next layer. With each layer of the network, the representation of the data becomes more abstract, although with increasing ability to distinguish different data classes [11]. The most common architectures in deep learning that are currently used for surgical applications are convolutional neural networks, recurrent neural networks, and residual neural networks.

APPLICATIVE FIELDS OF ARTIFICIAL INTELLIGENCE

The techniques, described previously, have been used to great effect within several applicative subfields of AI. Two of the most common in medicine (and in surgery more specifically) are the fields of computer vision (CV) and natural language processing (NLP).

Computer vision

CV is, in its simplest explanation, machine understanding of images and videos [3]. It is a subfield of AI but also composed of other fields like signal processing, pattern recognition, and image processing (Fig. 3—although CV does not subsume reinforcement learning, this merely represents overlap in fields). It involves a machine integrating information from the pixels that make up an image, detecting objects within an image, and potentially even engaging in analysis of open spaces within an image. These elements together can result in advanced applications, such as an autonomous driving system, where the computer is able to identify open roads, pedestrians, traffic lights, and so forth. CV also has benefited greatly from deep learning techniques.

Most of the readily recognizable advances in CV have come from the fields of radiology and pathology, perhaps due to the readily available nature of digital images in both fields. CV also has demonstrated promise with screening applications in ophthalmology, such as through the automated detection of diabetic retinopathy, and dermatology, where automated recognition of benign versus malignant skin lesions has been described [12,13]. Outside of specific clinical areas, CV and deep learning have been used together to create tools to predict radiation and magnetic exposure to staff in procedures where

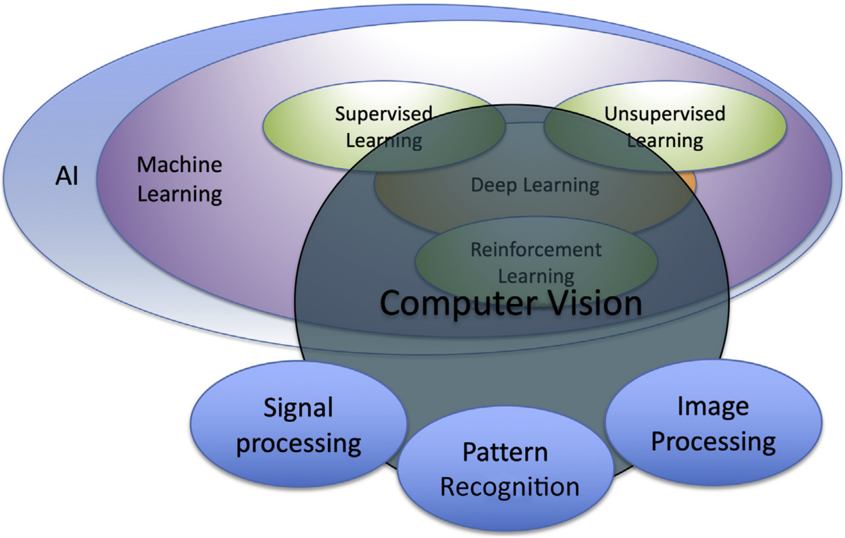

Fig. 3. Relationship of CV as a field to topics within AI. For illustration purposes, CV appears to subsume reinforcement learning but in reality reinforcement learning is one technique within and outside of CV as a field.

computed tomography, fluoroscopy, or magnetic resonance imaging may be used and combine such predictions with augmented reality, to provide feedback to the staff [14].

CV applications in surgery are increasing, however, as access to visual surgical data increases. With greater, cheaper storage capacities and more user-friendly laparoscopic, endoscopic, and robotic camera systems, many surgeons are choosing to record their operations for teaching, education, and research purposes. These applications are described later.

Natural language processing

NLP focuses on machine understanding of human language beyond identification of vocabulary (synonyms, antonyms, definitions, and so forth). Without NLP, computers are limited to reading machine languages or code (eg, C+, Java, and Visual Basic) to execute instructions based on explicitly programmed code that are compiled to yield an output. NLP allows machines to approximate the understanding of human language as it would be used in day-to-day life. It strives to achieve understanding of syntax and semantics to approximate meaning from phrases, sentences, or paragraphs [15].

NLP is perhaps most readily recognized in home assistant devices, such as Amazon Alexa (Amazon, Seattle, Washington) or Google Home (Alphabet, Mountain View, California). Analogous functions are found in digital platforms used for operative dictation (eg, Dragon software [Nuance, Burlington, Massachusetts]) Beyond the provider-facing functions, such as dictation, NLP is utilized heavily in the analysis and utilization of data within the electronic medical record. Because NLP can be used to analyze some forms of human language, unstructured free text, such as radiology reports, progress reports, and operative notes, can be analyzed and structured in an automated manner. As examples, it can be utilized to assess for sentiment in patient notes for the prediction of patient health status, to analyze records for risk prediction in cancer patients, and to detect surgical site infection from providers' notes [16–18].

SPECIFIC APPLICATIONS OF ARTIFICIAL INTELLIGENCE IN SURGERY

Preoperative risk prediction

Surgery is a controlled insult on the human physiology that is not without risks: 20% of surgeries have complications [19]. Therefore, prior to an operation, surgeons and patients have always asked, "What are the risks associated with this operation?" Ideally, risk prediction would both guide patient-centered decisions to evaluate both operative candidacy and predict possible postoperative complications.

Many risk calculators and decision algorithms exist on the market. Because adequate cardiac function is of utmost importance, the most prominent calculators assess and predict risk of major adverse cardiac events (MACE). Examples include the Revised Cardiac Risk Index (RCRI) and Gupta Perioperative Risk for Myocardial Infarction or Cardiac Arrest (MICA) [19–21]. Unfortunately, these models often underperform. The POISE trials showed a MACE rate of 6.9% in patients rather than the RCRI-predicted rate of 1% to 2.4% [22]. Outside of risk calculators, subjective patient-reported measures of cardiac functional capacity, such as metabolic equivalents, also tend to under-triage patients. One study found that subjects report their cardiac capacity, with a sensitivity of only 19% [23].

Other risk calculators incorporate more than cardiac factors alone to try to predict risk. Most famously, the American Society of Anesthesiologists (ASA) have created a classification system to predict risk [24]. Just like metabolic equivalents, however, the score has built-in subjectiveness, which leads to classification variability. For example, a survey of board-certified anesthesiologists found widespread variability in classification for more than a third of patients [25].

More recent efforts have attempted to solve the prior predictors' shortcomings through the use of objective big data to address model underperformance. For example, the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) released a risk calculator. Their calculator used information from 393 hospitals with approximately 1.5 million patients to create a generalized linear mixed model to predict risk of mortality and various complications. This model had good performance, with C statistics of 0.944 and 0.816 for mortality and morbidity, respectively [26].

Until recently, a majority of risk calculators have used traditional linear and additive models for risk prediction. Recent advances utilize ML methods for better approximation of the nonlinearity of patient risk factors. Researchers at Duke University trialed 3 different ML methods on their single-institution database of 100,000 patients: least absolute shrinkage and selection operator (lasso) penalized logistic regression, random forest models, and extreme gradient boosted trees [27]. Comparing the 3 methods, they found that lasso performed best for more than 8 of 14 outcomes whereas extreme gradient boosted trees excelled in 5 of 14 outcomes with areas under the curve ranging from 0.747 to 0.924. With their algorithms, they created an online calculator with 9 input data fields that out-performed the ACS-NSQIP calculator in post-operative mortality and morbidity prediction across-the-board for a random sample of 75 patients.

Similar work has come out of the University of Florida with their MySurgeryRisk score. They used their electronic medical record (EMR) data to create risk-prediction scores using ML techniques, such as random forests. Their risk-prediction was particularly patient-tailored because they linked training data to census data tied to zip codes and to surgeon-specific outcomes. Beyond just the creation of a risk-calculator, they also created interfaces for seamless EMR integration, not only so that risk-prediction happened in real time but also their models underwent continuous learning and tuning from physician-feedback [28].

With all these advances, the elephant-in-the-room question, "Do these scores actually improve upon current standards?" remains. A group from the Massachusetts General Hospital created another risk prediction calculator called Predictive OpTimal Trees in Emergency Surgery Risk (POTTER). They used the ML technique of optimal classification trees trained on 7 years of ACS-NSQIP data, which they packaged in a smartphone application for ease of use and deployment. Unlike the other scores, discussed previously, they compared their technique to multiple scores, including the ASA's and ACS-NSQIP, with superior performance across an entire year of patients from the ACS-NSQIP database [29].

Preoperative risk calculation has evolved over the past decades. Medical practitioners no longer are limited to relying on tried-and-true but subjective methods like clinical gestalt or classifications and instead can combine the ever-increasing big data from EMR with ML algorithms for objective, and increasingly accurate, predictions of patient outcomes. With continuous EMR integration and even deployment to smartphones, this wealth of information is immediately available and creates the promise that 1 day, patients may be able to be exactly answered when they ask, "What are the risks of this surgery?"

Intraoperative video analysis

AI technology has advanced significantly for the preoperative phase, but there have been few forays into intraoperative AI deployment. AI, through CV, allows computers to comprehend visual cues and, therefore, interact with the world in real time. With sufficient training and incorporation of thousands of operations, an AI model could guide surgeons, in real time, just as if they had a world expert in surgery looking over their shoulder. It already is known that experience matters, with an inverse relationship between a surgeons' case volumes and their patients' mortality [30]. Analysis of a surgeon's ability on visual cues alone is even predictive of a surgeon's rate of complications [31]. If a CV system could guide surgeons and take their performance from the bottom quartile to the top quartile, patients would receive immediate improvement in their care. Although it is known that skill does contribute to performance differences, approximately 70% of cases have near-miss events, two thirds of which need additional intervention to fix, which is something a CV model could warn the operator about and prevent from happening in the first place [32].

With the promise of a safer operating room, a few groups across the world have tried to tackle the difficult problem of teaching a computer to see and think like a high-level surgeon. CV is still quite new, with accurate image recognition possible only since 2012, so applications of CV in surgery are nascent [32,33]. The initial work has involved analysis of laparoscopic cases, given the ease of video acquisition and camera stability. In particular, groups have worked on identification of surgical phase with good accuracy across cholecystectomy (86.7%), sleeve gastrectomy (85.6%), and sigmoidectomy (91.9%) [34–36]. Additional applications of such technology have been investigated for its potential impact on improving operating room workflow and logistics, such as through the prediction of remaining operative time from intraoperative video alone [37].

Knowing that accurate phase recognition is possible, next steps will include development of intraoperative decision support. For example, likely applications include guidance for port placement, confirmation that a critical portion of the case has successfully been obtained (eg, the critical view of safety in cholecystectomy), and, in the more distant future, maybe even a real-time intraoperative global positioning system to guide surgeons in their dissection. As the AI models train with an increasing number of cases, they soon will develop an unparalleled surgical knowledge–a collective surgical consciousness–that will help any surgeon, anywhere, to deliver optimal intraoperative care to their patients [3].

There are, however, key advances that must be made. An important, early advance in the process of translating CV to the operating room is the establishment of clear labels for operative videos. Hashimoto and colleagues [34] (2019) demonstrated that surgeons, even within the same institution, can differ in their conceptualization of the boundaries of the steps of an operation. That is, when does 1 step of an operation end and the next begin? As described previously, for supervised learning, defining a gold standard or ground truth is important to be able to train a model to recognize aspects of surgical video [38]. Efforts currently are under way through the Society of American Gastrointestinal and Endoscopic Surgeons to convene an international consensus on guidelines for annotating operative video for the purposes of ML and CV research.

Electronic Health Records

Electronic medical records (EMRs) largely have benefited from some applications of AI, especially through NLP and other ML algorithms for risk calculation and resource management, as described in this article. Another potential great application of extracting data from EMR through AI techniques is the integration of the preoperative knowledge with the intraoperative events and postoperative outcomes of each individual. Furthermore, analyzing and linking each individual surgeon's intraoperative decisions and conducts and their postoperative data with populational data from several other surgeons would allow the machine to make inferences, predictions, and recommendations based on a collective knowledge and exponentially enhance cognitive capabilities by providing experience that would have taken several years to decades to achieve and then apply to all and each individual patients.

REGULATORY AND LEGAL CONSIDERATIONS

Like in any other industry, AI eventually will permeate almost all aspects of the surgical practice, and several regulatory bodies will be dictating how hospitals and surgeons can utilize this type of technology. And because surgical AI is still nascent, it is difficult to predict the unique future rules and regulations regarding its utilization, but it can be anticipated that they will be exercised at different levels, such as federal and state medical boards and individual hospitals and societies.

Legal considerations also likely will play a major role in surgical AI, because large amounts of data acquisition, especially video data, are paramount for the scalability and sustainability of the use of AI in surgery; therefore, broad policies governing data acquisition, storage, sharing, and utilization should be designed and agreed by surgeons, lawmakers, ethicists, privacy officers, engineers, payers, insurers, and patients.

SURGEON'S ROLE

It is unequivocal that this transformation will occur, and it will be driven by many forces, including financial, regulatory, media, and societal; and surgeons need to be prepared not only to embrace it but also to be instrumental from its inception to implementation. It is the surgeon's responsibility to understand, embrace, and advocate for the necessary changes in this field for this transformation to occur in an orderly and controlled manner, allowing interoperability, scalability, and sustainability for the years to come.

Surgeons should work closely with engineers and data scientists to develop the technologies at the very early stages and work with academic societies, medical boards, and hospital administration to assess and safely implement the nascent technology in a comprehensive and thoughtful fashion.

Furthermore, surgeons also have an important leadership role to educate other health care providers, hospital staff, industry, patients, medical students, population, and lawmakers on the benefits and potential pitfalls of this type of technology.

SUMMARY

In summary, AI as applied to surgery is early in its development. Although significant advances are being made in AI, these advances are focused on narrow applications of the technology to specific problems within surgery. The field is very much in a phase of discovery and development, and a critical appraisal of new publications, software, and devices is necessary to appropriately evaluate its impact on patient care and surgeon workflow. As with any new technology, a healthy measure of skepticism is necessary to guard against hype; however, data on potential applications of AI to surgery thus far have been promising.

Disclosures

D.A. Hashimoto: consultant for Verily Life Sciences, Johnson and Johnson Institute, and Mosaic Research Management; advisory board for Worrell Inc.; and research support from Olympus Corporation. T.M. Ward: research support from Olympus Corporation. O.R. Meireles: consultant for Olympus America and Medtronic and research support from Olympus Corporation.

References

[1] Bellman R. An introduction to artificial intelligence: can computers think? Thomson Course Technology; 1978.
[2] Administration USF&. D, U.S. Food & Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Case Med Res 2018; https://doi.org/10.31525/fda2-ucm604357.htm.
[3] Hashimoto DA, Rosman G, Rus D, et al. Artificial Intelligence in Surgery: Promises and Perils. Ann Surg 2018;268(1):70–6.
[4] Hashimoto DA, Witkowski E, Gao L, et al. Anesthesiology 2019;1; https://doi.org/10.1097/aln.0000000000002960.
[5] Avanzolini G, Barbini P, Gnudi G. Unsupervised learning and discriminant analysis applied to identification of high risk postoperative cardiac patients. Int J Biomed Comput 1990;25(2–3):207–21.
[6] DiPietro R, Hager GD. Unsupervised learning for surgical motion by learning to predict the future. In: Medical image computing and computer assisted intervention – MICCAI 2018. 2018. 2018. p. 281–8; https://doi.org/10.1007/978-3-030-00937-3_33.
[7] Skinner BF. The behavior of organisms: an experimental analysis. B. F. Skinner Foundation; 1990.
[8] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge. Nature 2017;550(7676):354–9.
[9] Hebb DO. The organization of behavior. New York: Psychology Press; 2005; https://doi.org/10.4324/9781410612403.
[10] Natarajan P, Frenzel JC, Smaltz DH. Demystifying big data and machine learning for healthcare. Boca Raton (FL): CRC Press; 2017; https://doi.org/10.1201/9781315389325.
[11] Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019;25(1):24–9.
[12] Grzybowski A, Brona P, Lim G, et al. Artificial intelligence for diabetic retinopathy screening: a review. Eye 2019; https://doi.org/10.1038/s41433-019-0566-0.
[13] Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115–8.
[14] Rodas NL, Padoy N. Augmented reality for reducing intraoperative radiation exposure to patients and clinicians during X-ray guided procedures. In: Peters TM, Linte CA, Yaniv Z, et al, editors. Mixed and augmented reality in medicine. Boca Raton (FL): CRC Press; 2018. p. 217–29; https://doi.org/10.1201/9781315157702-15.

[15] Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011;18(5):544–51.
[16] Hughes KS, Zhou J, Bao Y, et al. Natural language processing to facilitate breast cancer research and management. Breast J 2020;26(1):92–9.
[17] Zunic A, Corcoran P, Spasic I. Sentiment analysis in health and well-being: systematic review. JMIR Med Inform 2020;8(1):e16023.
[18] Shen F, Larson DW, Naessens JM, et al. Detection of surgical site infection utilizing automated feature generation in clinical notes. Int J Healthc Inf Syst Inform 2019;3(3):267–82.
[19] Healey MA, Shackford SR, Osler TM, et al. Complications in surgical patients. Arch Surg 2002;137(5):611–7 [discussion: 617–8].
[20] Lee TH, Marcantonio ER, Mangione CM, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation 1999;100(10):1043–9.
[21] Gupta PK, Gupta H, Sundaram A, et al. Development and validation of a risk calculator for prediction of cardiac risk after surgery. Circulation 2011;124(4):381–7.
[22] POISE Study Group, Devereaux PJ, Yang H, et al. Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): a randomised controlled trial. Lancet 2008;371(9627):1839–47.
[23] Wijeysundera DN, Pearse RM, Shulman MA, et al. Assessment of functional capacity before major non-cardiac surgery: an international, prospective cohort study. Lancet 2018;391(10140):2631–40.
[24] Wolters U, Wolf T, Stützer H, et al. ASA classification and perioperative variables as predictors of postoperative outcome. Br J Anaesth 1996;77(2):217–22.
[25] Owens WD, Felts JA, Spitznagel EL Jr. ASA physical status classifications: a study of consistency of ratings. Anesthesiology 1978;49(4):239–43.
[26] Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg 2013;217(5):833–42.e1-3.
[27] Corey KM, Kashyap S, Lorenzi E, et al. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data (Pythia): A retrospective, single-site study. PLoS Med 2018;15(11):e1002701.
[28] Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. MySurgeryRisk: development and validation of a machine-learning risk algorithm for major complications and death after surgery. Ann Surg 2019;269(4):652–62.
[29] Bertsimas D, Dunn J, Velmahos GC, et al. Surgical risk is not linear. Ann Surg 2018;268(4): 574–83.
[30] Birkmeyer JD, Stukel TA, Siewers AE, et al. Surgeon volume and operative mortality in the United States. ACC Curr J Rev 2004;13(2):59.
[31] Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med 2013;369(15):1434–42.
[32] Bonrath EM, Gordon LE, Grantcharov TP. Characterising ''near miss'' events in complex laparoscopic surgery through video analysis. BMJ Qual Saf 2015;24(8):516–21.
[33] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90.
[34] Hashimoto DA, Rosman G, Witkowski ER, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. Ann Surg 2019;270(3):414–21.
[35] Kannan S, Yengera G, Mutter D, et al. Future-state predicting LSTM for early surgery type recognition. IEEE Trans Med Imaging 2019; https://doi.org/10.1109/TMI.2019.2931158.
[36] Kitaguchi D, Takeshita N, Matsuzaki H, et al. Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based

deep learning approach. Surg Endosc 2019; https://doi.org/10.1007/s00464-019-07281-0.
[37] Twinanda AP, Yengera G, Mutter D, et al. RSDNet: learning to predict remaining surgery duration from laparoscopic videos without manual annotations. IEEE Trans Med Imaging 2019;38(4):1069–78.
[38] 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; https://doi.org/10.1007/s00464-019-07149-3.