

ANESTHESIOLOGY

Artificial Intelligence in Anesthesiology

Current Techniques, Clinical Applications, and Limitations

Daniel A. Hashimoto, M.D., M.S., Elan Witkowski, M.D., M.P.H., Lei Gao, M.D., Ozanan Meireles, M.D., Guy Rosman, Ph.D.

ANESTHESIOLOGY 2020; 132:379–94

Artificial intelligence has been defined as the study of algorithms that give machines the ability to reason and perform functions such as problem-solving, object and word recognition, inference of world states, and decision-making.¹ Although artificial intelligence is often thought of as relating exclusively to computers or robots, its roots are found across multiple fields, including philosophy, psychology, linguistics, and statistics. Thus, artificial intelligence can look back to visionaries across those fields, such as Charles Babbage, Alan Turing, Claude Shannon, Richard Bellman, and Marvin Minsky, who helped to provide the foundation for many of the modern elements of artificial intelligence.² Furthermore, major advances in computer science, such as hardware-based improvements in processing and storage, have enabled the base technologies required for the advent of artificial intelligence.

Artificial intelligence has been applied to various aspects of medicine, ranging from largely diagnostic applications in radiology³ and pathology⁴ to more therapeutic and interventional applications in cardiology⁵ and surgery.⁶ In April 2018, the U.S. Food and Drug Administration approved the first software system that uses artificial intelligence—a program that assists in the diagnosis of diabetic retinopathy through the analysis of images of the fundus.⁷ As the development and application of artificial intelligence technologies in medicine continues to grow, 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, more cost-effective care.

This article is featured in "This Month in Anesthesiology," page 1A. Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text of this article on the Journal's Web site (www.anesthesiology.org).

Submitted for publication November 26, 2018. Accepted for publication July 29, 2019. Published online first on September 12, 2019. From the Surgical Artificial Intelligence and Innovation Laboratory (D.A.H., E.W., O.M., G.R.) and Department of Anesthesia, Critical Care, and Pain Medicine (L.G.), Massachusetts General Hospital, Boston, Massachusetts; Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts (G.R.).

Copyright © 2020, the American Society of Anesthesiologists, Inc. All Rights Reserved. *Anesthesiology* 2020; 132:379–94. DOI: 10.1097/ALN.0000000000002960

ABSTRACT

Artificial intelligence has been advancing in fields including anesthesiology. This scoping review of the intersection of artificial intelligence and anesthesia research identified and summarized six themes of applications of artificial intelligence in anesthesiology: (1) depth of anesthesia monitoring, (2) control of anesthesia, (3) event and risk prediction, (4) ultrasound guidance, (5) pain management, and (6) operating room logistics. Based on papers identified in the review, several topics within artificial intelligence were described and summarized: (1) machine learning (including supervised, unsupervised, and reinforcement learning), (2) techniques in artificial intelligence (*e.g.*, classical machine learning, neural networks and deep learning, Bayesian methods), and (3) major applied fields in artificial intelligence.

The implications of artificial intelligence for the practicing anesthesiologist are discussed as are its limitations and the role of clinicians in further developing artificial intelligence for use in clinical care. Artificial intelligence has the potential to impact the practice of anesthesiology in aspects ranging from perioperative support to critical care delivery to outpatient pain management.

(*ANESTHESIOLOGY* 2020; 132:379–94)

Anesthesiology as a field is well positioned to potentially benefit from advances in artificial intelligence as it touches on multiple elements of clinical care, including perioperative and intensive care, pain management, and drug delivery and discovery. We conducted a scoping review of the literature at the intersection of artificial intelligence and anesthesia with the goal of identifying techniques from the field of artificial intelligence that are being used in anesthesia research and their applications to the clinical practice of anesthesiology.

Materials and Methods

The Medline, Embase, Web of Science, and IEEE Xplore databases were searched using combinations of the following keywords: machine learning, artificial intelligence, neural networks, anesthesiology, and anesthesia. To be included, papers had to be focused on the design or application of artificial intelligence-based algorithms in the practice of anesthesia, including preoperative, intraoperative, postoperative, and surgical critical care as well as pain management. All English-language papers from 1946 to September 30, 2018 were eligible. Peer-reviewed, published literature, including narrative review papers, were eligible for inclusion as were peer-reviewed conference proceeding papers. Editorials, letters to the editor, and abstracts were excluded as were any studies involving animals, fewer than 10 patients, or simulated data only. Papers that involved monitoring of

patient parameters broadly but were not specific to the practice of anesthesia (e.g., general ward-based monitoring and alarms, sleep study analysis) were excluded.

Four reviewers screened articles for inclusion or exclusion using Covidence (Melbourne, Australia). Each article was screened by two independent reviewers. A third reviewer would mediate any disagreement between two screeners. Reference lists of included papers were hand-searched by one reviewer and included if the inclusion criteria were met.

Emphasis was placed on extracting themes relating to applications of artificial intelligence. Although specifics regarding the numerous algorithms that can be used in artificial intelligence studies and applications were outside the scope of this review, common techniques uncovered in the literature were reviewed as examples.

Results

A total of 2,171 titles were identified from the database search, and 1,368 abstracts were eligible for screening. Of these, 394 full-text studies were assessed for inclusion, and 173 manuscripts were included in the final analysis (fig. 1).

Based on the 173 included full-text articles, we identified the following categories of studies: (1) depth of anesthesia monitoring, (2) control of anesthesia, (3) event and risk prediction, (4) ultrasound guidance, (5) pain management, and (6) operating room logistics. Seven articles (n = 7) were broad

review papers that covered multiple categories. Included articles and their categorization are listed in Supplemental Digital Content (<http://links.lww.com/ALN/C51>). From these categories, we identified the major topics of artificial intelligence that were captured in this search.

The predominant focus across most of these studies has been to investigate potential ways that artificial intelligence can benefit the clinical practice of anesthesiology not through the replacement of the clinician but through augmentation of the anesthesiologist’s workflow, decision-making, and other elements of clinical care. Thus, although artificial intelligence is an expansive field, the results of this review demonstrated the literature’s focus on machine learning and its applications to clinical care. To structure the findings of the scoping review, we have presented below the following: machine learning and the types of learning algorithms, major topics in artificial intelligence, artificial intelligence techniques, and the six aforementioned categories of applications of artificial intelligence to anesthesiology that were found in the literature.

Machine Learning and Learning Algorithms in Artificial Intelligence

Although different taxonomies of artificial intelligence have been previously described, common to many of these is the categorization of machine learning as one of the major

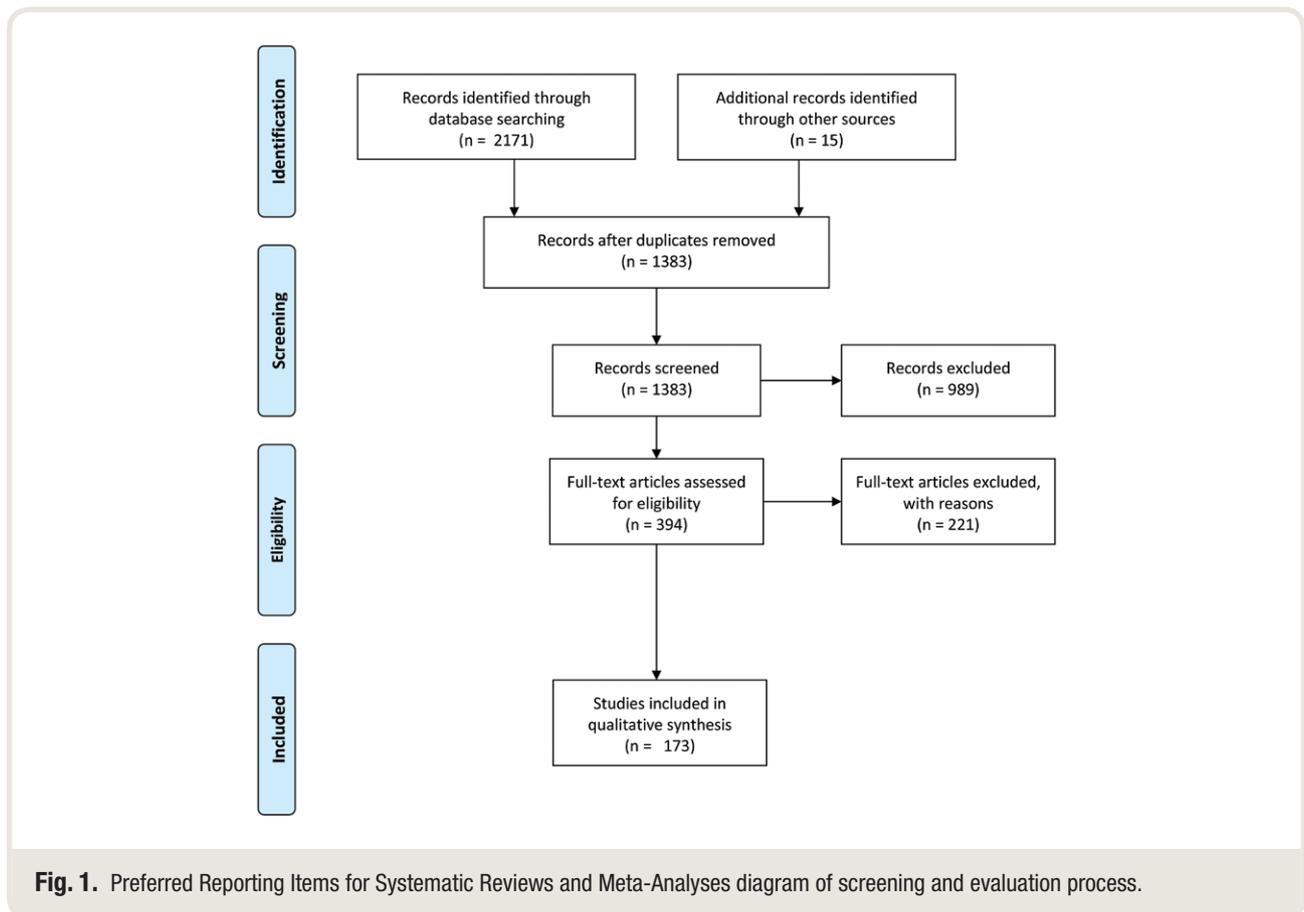


Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram of screening and evaluation process.

subfields of artificial intelligence. Traditional computer programs are programmed with explicit instructions to elicit certain behaviors from a machine based on specific inputs (e.g., the primary function of a word processing program is to display the text input by the user). Machine learning, on the other hand, allows for programs to learn from and react to data without explicit programming. Data that can be analyzed through machine learning are broad and include, but are not limited to, numerical data, images, text, and speech or sound. A common manner with which to conceptualize machine learning is to consider the type of learning algorithm used to solve a problem: supervised learning, unsupervised learning, and reinforcement learning.⁸ Each of these types of learning algorithms has a range of techniques that can be applied and will be covered in the next section.

Supervised learning is a task-driven process by which an algorithm(s) is trained to predict a prespecified output, such as identifying a stop sign or recognizing a cat in a photograph. Supervised learning requires both a training dataset and a test dataset. The training dataset allows the machine to analyze and learn the associations between an input and desired output while the test dataset allows for the assessment of the performance of the algorithm on new data. In many studies, one large dataset is subdivided into a training set and a test set (often 70% of the data for training, 30% for testing).

For example, Kendale *et al.*⁹ conducted a supervised learning study on electronic health record data with the goal of identifying patients who experienced postinduction hypotension (mean arterial pressure [MAP] less than 55 mmHg). The training dataset included 70% of the patients and variables such as American Society of Anesthesiologists (ASA) physical status, age, body mass index, comorbidities, and medications as well as the blood pressures of the patients. The various algorithms used by Kendale *et al.*⁹ could then analyze the training dataset to learn which variables were predictive of postinduction hypotension. The test dataset was then analyzed to assess how accurately the algorithm could predict postinduction hypotension in the remaining 30% of patients.⁹ Some studies use external validation (*i.e.*, the use of a separate dataset) to assess the generalizability of the algorithm to other data sources.¹⁰ A study using electronic health record data from one hospital in their training and testing datasets may then incorporate electronic health record data from a separate hospital to further validate the performance of their algorithm on a different data source, demonstrating generalizability to other populations.

Unsupervised learning refers to algorithms identifying patterns or structure within a dataset. This can be useful to find novel ways of classifying patients, drugs, or other groups. Bisgin *et al.*¹¹ used unsupervised learning techniques to mine data from the Food and Drug Administration drug labels to identify the major topics (e.g., specific adverse events, therapeutic application) into which drugs could be

automatically classified for hypothesis generation for future research.¹¹ Similarly, unsupervised learning can be used to identify patients who could most benefit from certain drugs, such as asthmatics who would benefit most from glucocorticoid therapy, based on genomic analysis.¹²

Reinforcement Learning refers to the process by which an algorithm is asked to attempt a certain task (e.g., deliver inhalational anesthesia to a patient, drive a car) and to learn from its subsequent mistakes and successes.¹³ A biologic analogy to reinforcement learning is operant conditioning, where the classical example is of a rat taught to push a lever through the use of food-based reward. However, reinforcement learning problems today are more sophisticated. For example, Padmanabhan *et al.*¹⁴ used reinforcement learning to develop an anesthesia controller that used feedback from a patient's bispectral index (BIS) and MAP to control the infusion rates of Propofol (in a simulated patient model). In this scenario, achieving BIS and MAP values within a set range result in a reward to the algorithm whereas values outside the range result in errors that prompt the algorithm to perform further fine-tuning.

Machine learning problems are often divided into those that require classification (dividing data into discrete groups) and those that require regression (modeling data to better understand the relationship between two or more continuous variables with the potential for prediction). A frequent example of classification is image recognition (e.g., recognizing a cat *vs.* a dog), whereas a frequent example of regression is prediction (e.g., predict house prices from preexisting data on square footage).¹⁵ Supervised, unsupervised, and reinforcement learning approaches can each be used to tackle problems of classification and regression depending on the nature of the question and the type of data available.

Techniques in Artificial Intelligence

There are various techniques and models within each of the three approaches to machine learning described above. Although a detailed description of the specific methods and algorithms used in machine learning are outside the scope of this review, it can be useful to have an introductory familiarity with basic concepts of the more popular techniques used in artificial intelligence research.

Fuzzy Logic – A Historical Perspective. Fuzzy set theory and fuzzy logic were first described in 1965.¹⁶ Although fuzzy logic in and of itself is not necessarily artificial intelligence, it has been used within other frameworks to facilitate artificial intelligence-based functions. Standard logic allows only for the concepts of true (a numerical value of 1.0) and false (a numerical value of 0.0), but fuzzy logic allows for partial truth (*i.e.*, a numerical value between 0.0 and 1.0). A comparison may be made to probability theory where the probability of a statement being true is evaluated (e.g., “a pancreaticoduodenectomy will be scheduled tomorrow”)

versus the degree to which a statement is true (e.g., “there is an 80% chance of a pancreaticoduodenectomy being scheduled tomorrow”). It is meant to resemble human methods of decision making with vague or imprecise information.

Fuzzy logic often uses rule-based systems (e.g., if...then systems) that have been largely used in control systems where precise mathematical functions do not accurately model phenomena. For example, an anesthesia monitor to detect hypovolemia was developed using fuzzy logic to approximate the presence of mild, moderate, and severe hypovolemia based on normalized values of heart rate (HR), blood pressure, and pulse volume that were divided into categories of mild, moderate, and severe. The monitor used rules built on fuzzy logic such as “If (electrocardiogram-HR is mild) and (blood pressure is mild) and (pulse volume is severe), then (hypovolemia is moderate).”¹⁷

The development of such rules requires expert human input to determine an appropriate rule set for the machine to follow, and early studies in fuzzy logic and other adaptive control mechanisms help set the foundation for exploring more modern approaches to imprecise information or incomplete data. More recent research in the field has used artificial intelligence approaches to better evaluate and use data to trigger the rule functions of fuzzy systems. Thus, although research in fuzzy logic systems is ongoing (especially in control systems for applications such as medication delivery), progress in artificial intelligence research has focused on using more data-driven techniques in machine learning to achieve goals first explored by researchers working with fuzzy logic.

Classical Machine Learning. Machine learning uses features, or properties within the data, to perform its tasks. To analogize to an example in statistical analysis, features would be analogous to independent variables in a logistic regression. In classical machine learning, the features are selected (often referred to as *hand-crafted*) by experts to help guide the algorithms in the analysis of complex data.

Decision tree learning is a type of supervised learning algorithm that can be used to perform either classification (classification trees) or regression tasks (regression trees). As its name implies, this set of techniques uses flowchart-like tree models with multiple branch points to determine a target value or classification from an input. Each node within a tree has an assigned value, and the final node represents an endpoint and the cumulative probability of arriving at that point based on the preceding decisions (fig. 2). Hu *et al.*¹⁸ used decision trees to predict total patient-controlled analgesia (PCA) consumption from features such as patient demographics, vital signs and aspects of their medical history, surgery type, and PCA doses delivered with the promise of using such approaches to optimize PCA dosing regimens.¹⁸

The *k*-nearest neighbor algorithms are a group of supervised learning algorithms that assesses training data

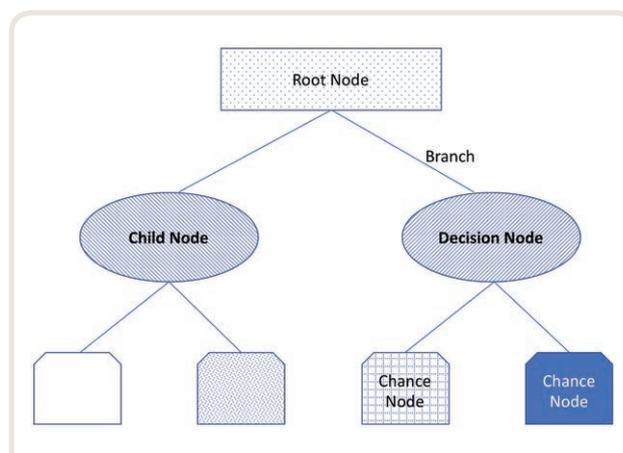


Fig. 2. An illustrative example of a decision node. Several terminologies can be used to describe decision trees. The root node is the start of the tree, and branches connect nodes. A child node is any node that has been split from a previous node, whereas a decision node is any node that allows two or more options to follow it. A chance node is any node that may represent uncertainty.

geometrically and then determine whether additional input data falls into a category based on the training examples plotted nearest to it (based on Euclidian distance). Based on the specific approach used, this may be based on a single nearest point (1-nearest neighbor) or the weight of a group of points (*k*-nearest neighbor).¹⁹ Support vector machines are another type of supervised learning algorithm that can also be used for both classification and regression. Support vector machines map training data in space and then use hyperplanes to optimally separate the data into representative categories or clusters (fig. 3). New data are then classified based on their location in the space relative to the hyperplane.²⁰

Neural Networks and Deep Learning. One of the most popular methods today for performing work in machine learning is the use of neural networks. Neural networks are inspired by biologic nervous systems and process signals in layers of computational units (neurons).²¹ Each network is made up of an input layer of neurons comprised of features that describe the data, at least one hidden layer of neurons that conducts different mathematical transformations on the input features, and an output layer that yields a result (fig. 4). Between each layer are multiple connections between neurons that are parameterized to different weights depending on the input-output maps. Thus, neural networks are a framework within which different machine learning algorithms can work to achieve a particular task (e.g., image recognition, data classification).

Modern neural network architecture has expanded to allow for deep learning, neural networks that use many layers to learn more complex patterns than those that

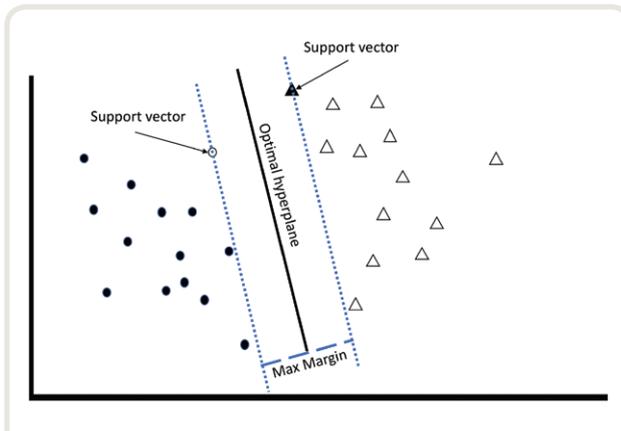


Fig. 3. An illustrative example of support vector machines. The goal of the support vector machines algorithm is to find the hyperplane that maximizes the separation of features. The *solid black line* represents the optimal hyperplane, whereas the *dotted lines* represent the planes running through the support vectors. The *empty circle* and the *solid triangle* represent support vectors—the data points from each cluster that represent the closest points to the optimal hyperplane. The *dashed line* represents the maximum margin between the support vectors.

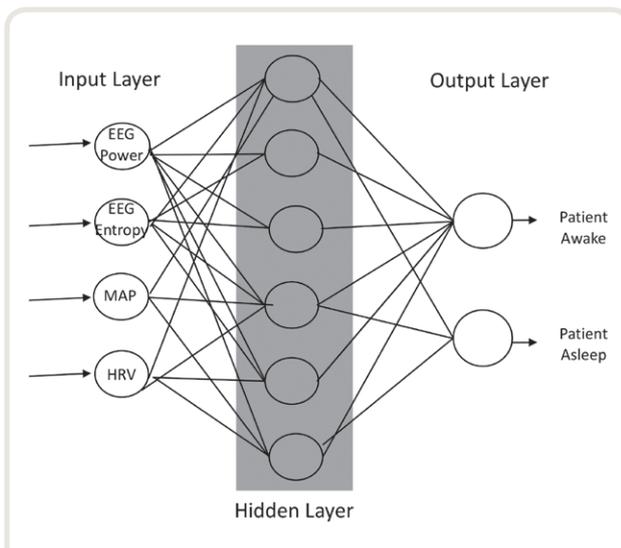


Fig. 4. An illustrative example of a three-layer neural network. The input layer provides features such as electroencephalogram (EEG) power and entropy, the patient's mean arterial pressure (MAP), and the patient's heart rate variability (HRV) to the network. A hidden layer transforms inputs into features usable by the network. The output layer transforms the hidden layer's activations into an interpretable output (*e.g.*, patient awake vs. asleep).

are discernible from simple one- or two-layer networks. Subtypes of deep learning networks that one may encounter are convolutional neural networks, which can process data composed of multiple arrays, and recurrent neural networks, which are better designed to analyze sequential data

(*e.g.*, speech).²² Unlike in classical machine learning where features are hand-crafted, deep learning self-learns features based on the data itself. It analyzes all available features within the training set to determine which features allow for the optimal achievement of the deep neural network's given task (*e.g.*, object recognition from an image). Thus, deep learning is potentially a powerful tool with which to analyze very large datasets where hand-crafted feature selection may be inadequate and/or inefficient in achieving an effective result.

Given their flexibility in analyzing various types of data, neural networks are a technique that have now been applied to other subfields of artificial intelligence, including natural language processing and computer vision. Within the field of anesthesia, multiple examples of the applications of neural networks exist, including depth of anesthesia monitoring and control of anesthesia delivery as described further below.

Bayesian Methods. Bayes' theorem provides a description of probability of an event based on previous knowledge or data about factors that may influence that event. In many studies in the medical literature, a frequentist approach to statistics is applied, wherein hypothesis testing occurs based on the frequency of events that occur in a given sample of data as a representation of the study population of interest. A Bayesian approach differs by using the known previous probability distribution of an event along with the probability distribution represented in a given dataset.²³

Bayes' theorem provides the foundation for many techniques in artificial intelligence as it allows for both modeling of uncertainty and updating or learning repeatedly as new data is made available.²⁴ Bayesian techniques have been incorporated into many common tasks such as spam filtering, financial modeling, and evaluation of clinical tests.^{25,26} Although delving into the specific methodologies of Bayesian methods is outside the scope of this review, Bayesian methods that are being used with increased frequency in the medical literature include Bayesian networks, hidden Markov models, and Kalman filters.²⁷

Applied Fields in Artificial Intelligence

Two other topics identified by the search as relevant to anesthesiology were natural language processing and computer vision, and these topics are both applications of and subfields within artificial intelligence that use machine learning.

Natural Language Processing. Natural language processing is a subfield of artificial intelligence that focuses on machine understanding of human language. Before the advent of natural language processing, computers were limited to reading machine languages or code (*e.g.*, C++, JAVA, Visual Basic). Instructions programed in code are compiled by a computer to process a set of instructions to yield a desired output. With natural language processing, machines can strive toward the understanding of language that is used naturally by

humans. Natural language processing, however, is not simply recognizing letters that construct a word and then matching them to a definition. It strives to achieve understanding of syntax and semantics to approximate meaning from phrases, sentences, or paragraphs.²⁸

In medicine, the most natural application of natural language processing is to automated analysis of electronic health record data. Although the move to electronic health records has shifted a considerable amount of documentation to checkboxes, dropdown menus, and prepopulated fields, free text entry remains a critical component of clinician documentation, allowing clinicians to communicate with one another and to document not just important findings but also the thought process behind medical decisions. Natural language processing is currently being used to help extract information from free text fields to build more structured databases that can be further analyzed to identify surgical candidates, assess for adverse events, or to facilitate billing.^{29–31}

Computer Vision. With the growing attention on autonomous vehicles, computer vision has become one of the most widely recognized subfields of artificial intelligence. Computer vision refers to machine understanding of images, videos, and other visual data (e.g., computed tomography). Automated acquisition, processing, and analysis of images are part of the computer vision process that leads to understanding of a scene. Color, shape, texture, contour, and focus are just some of the different elements that can be detected and analyzed by computer vision systems. While object detection and recognition (e.g., identifying another car on the road) have been popularized in the lay press, computer vision encompasses the field of work that includes research on describing the visual world in numerical or symbolic form to allow for interpretation of images for subsequent action (e.g., determining that the car on the road ahead of your will be slowing down at the stop sign, prompting the need to slow your vehicle as well).

In medicine, the application of computer vision to pathology and radiology have led to systems capable of assisting clinicians in reducing error rates in diagnosis by identifying areas on slides and x-rays that have a high probability of demonstrating pathology.^{32,33} Furthermore, computer vision has been used to automatically identify and segment steps of laparoscopic surgery, suggesting that context awareness is possible with computer vision systems.³⁴ In anesthesiology, computer vision has largely been applied to the automated analysis of ultrasound images to assist with identification of structures during procedures.^{35,36}

Applications of Artificial Intelligence in Anesthesiology

Depth of Anesthesia Monitoring. Use of artificial intelligence to improve depth of anesthesia monitoring was identified in 42 papers. The majority of these papers focused on use of the BIS (Medtronic, USA) or electroencephalography to assess

anesthetic depth. The attention to these two modalities of measurement was not surprising given research efforts in reducing the risk of intraoperative awareness and previous literature suggesting that low BIS and burst-suppression on electroencephalography during anesthesia may be associated with poorer outcomes.^{37,38} In addition, careful monitoring of MAP has also been noted in the literature, likely due to the association of low MAP with postoperative mortality.³⁹

Machine learning approaches are well-suited to analyze complex data streams such as electroencephalographies; thus, a range of electroencephalography-based signals was found to have been investigated to measure depth of anesthesia. Literature from the early 1990s described the use of neural networks to evaluate electroencephalography power spectra as a signal in discriminating awake *versus* anesthetized patients and identified the potential of specific frequency bands in assessing the effects of certain drugs.^{40,41} As use of index parameters of depth of anesthesia such as BIS increased in popularity, neural networks and other machine learning approaches were used to analyze electroencephalography data with the goal of approximating BIS through other electroencephalography parameters.^{42,43}

More recent papers have used artificial intelligence techniques and spectral analysis to more directly analyze electroencephalography signals to estimate the depth of anesthesia. Mirsadeghi *et al.*⁴⁴ studied 25 patients and compared the accuracy of their machine learning method of analyzing direct features from electroencephalography signals (e.g., power in different bands [delta, theta, alpha, beta and gamma], total power, spindle score, entropy, *etc.*) in identifying awake *versus* anesthetized patients against the BIS index. Their accuracy in using electroencephalography features was 88.4% while BIS index accuracy was 84.2%.⁴⁴ Similarly, Shalbaf *et al.*⁴⁵ used multiple features from electroencephalographies to classify awake *versus* anesthetized patients (as four possible states of awake, light, general, or deep anesthesia) during sevoflurane with 92.91% accuracy compared with the response entropy index which had an accuracy of 77.5%. This same algorithm demonstrated with generalization to Propofol and volatile anesthesia patients with 93% accuracy *versus* the BIS index's 87% accuracy.⁴⁵ These papers highlight the power of artificial intelligence techniques in creating models that can efficiently consider linear and non-linear data together to generate predictions that maximize the utility of each available variable.

Whereas BIS and electroencephalography were the subject of the majority of the papers identified, other clinical signals have been investigated as well. Zhang *et al.*⁴⁶ recorded mid-latency auditory evoked potentials from patients and used neural networks to assess the accuracy of these signals in determining when patients were awake (96.8% accuracy), receiving adequate anesthesia (86% accuracy), and emerging from general anesthesia (86.6% accuracy). In addition, clinical variables such as heart rate variability have been investigated to approximate sedation level as

typically measured by the Richmond Agitation Sedation Scale.⁴⁶ Related to depth of anesthesia monitoring, Ranta *et al.*⁴⁷ conducted a database study of 543 patients who had undergone general anesthesia with 6% reporting intraoperative awareness. Neural networks analyzed clinical data from these patients such as blood pressure, heart rate, and end-tidal carbon dioxide but did not use any electroencephalography data. The prediction probabilities for these networks in predicting awareness was, at best, 66% though it had a high specificity of 98%.⁴⁷ As described above about neural networks and deep learning, the strength of such approaches over classical regression techniques is a neural network's ability to engage in self-learning of the features in the dataset, selecting features that will best predict a target value (*e.g.*, awareness) rather than being fed the features thought to be most predictive by a human expert.

Control of Anesthesia Delivery. Thirty-three ($n = 33$) papers were identified that pertained to control systems for delivery of anesthesia, neuromuscular blockade, or other related drugs. Control systems in anesthesia were reviewed in detail by Dumont and Ansermino⁴⁸ with explanations of feedforward and feedback systems as well as multiple examples of different closed-loop systems.⁴⁸ As automated delivery of anesthesia also requires a determination of the depth of anesthesia by a machine, approaches to control require the measurement of clinical signs or surrogate markers of anesthetic depth. Thus, the evolution of control system research in anesthesia is evident in the various targets used to approximate depth of anesthesia.

In the 1990s, clinical signs and measurements such as blood pressure were used as signals to the control systems to determine how to regulate delivery of anesthetics.^{49,50} As new metrics to measure depth of anesthesia were developed, targets shifted to measuring values such as BIS. Early work used purely empirical approaches to tuning delivery of anesthesia to achieve a target BIS.⁵¹ As use of BIS became more widespread, researchers began to use more sophisticated fuzzy logic systems or reinforcement learning to achieve anesthetic control using BIS as a target measure.⁵²⁻⁵⁴ Our search did not return papers on the use of direct electroencephalography measures for anesthesia control.

Control systems that use machine learning have also been used to automate the delivery of neuromuscular blockade,^{55,56} and these systems have also incorporated forecasting of drug pharmacokinetics to further improve the control of infusions of paralytics.⁵⁷ Three papers were identified that described the use of artificial intelligence to achieve control of mechanical ventilation^{58,59} or to automate weaning from mechanical ventilation.⁶⁰

Event Prediction. Our search identified 53 papers that could loosely be classified as pertaining to event prediction. We refer to event prediction as any studies that engage in prediction of an effect or event (*e.g.*, complication, length

of stay, awareness, *etc.*), and these papers were categorized into those pertaining to operative (immediate preoperative assessment; $n = 26$ papers), postoperative ($n = 14$ papers), or critical care-related events ($n = 13$ papers).

For perioperative care risk prediction, various techniques in machine learning, neural networks, and fuzzy logic have all been applied. For example, neural networks were used to predict the hypnotic effect (as measured by BIS) of an induction bolus dose of propofol (sensitivity of 82.35%, specificity of 64.38% and an area under the curve of 0.755) and was found to exceed the average estimate of practicing anesthesiologists (sensitivity: 20.64%, specificity 92.51%, area under the curve of 0.5605).⁶¹ At the tail end of an operation, Nunes *et al.* (2005)⁶² compared neural networks and fuzzy models in their prediction of return of consciousness after general anesthesia (propofol + remifentanyl) in a small sample of 20 patients, finding mixed results for all models. Neural networks have also been used to predict the rate of recovery from neuromuscular blockade⁶³ and hypotensive episodes postinduction⁶⁴ or during spinal anesthesia,⁶⁵ while other machine learning approaches have been tested to automatically classify pre-operative patient acuity (*i.e.*, ASA status),⁶⁶ define difficult laryngoscopy findings,⁶⁷ identify respiratory depression during conscious sedation,⁶⁸ and to assist in decision-making for the optimal method of anesthesia in pediatric surgery.⁶⁹

With regard to specific event detection in both the operating room and the intensive care unit (ICU), Hatib *et al.*⁷⁰ used two large databases (total of 1,334 patients) of waveform data from arterial lines to develop and test a logistic regression model that could predict hypotension (*i.e.*, mean arterial pressure less than 65 mmHg sustained for 1 min) up to 15 min before its occurrence on an arterial line waveform.⁷⁰ Other ICU database studies have used machine learning models to predict morbidity,^{71,72} weaning from ventilation,⁷³ clinical deterioration,^{74,75} mortality,⁷⁶ or readmission⁷⁷ and to detect sepsis.⁷⁸

Machine learning approaches to critical care have not been limited to large database studies only. In a single-center randomized control trial comparing a machine learning alert system (using six vital sign parameters as features) *versus* an electronic health record-based alert system that used other criteria for the prediction of sepsis, the machine learning alert system outperformed Systemic Inflammatory Response Syndrome criteria, Sequential Organ Failure Assessment score, and quick Sequential Organ Failure Assessment score in the detection of sepsis. Its use resulted in a 20.6% decrease in average hospital length of stay and, more importantly, a 58% reduction in in-hospital mortality.⁷⁹

Ultrasound Guidance. Eleven ($n = 11$) papers were identified that described the use of artificial intelligence techniques to assist in the performance of ultrasound-based procedures, and neural networks were the most commonly employed method of achieving ultrasound image classification. Smistad *et al.*³⁶ used ultrasound images of the groin from 15

patients to train a convolutional neural network to identify the femoral artery or vein while distinguishing it from other potentially similar appearing ultrasound images such as muscle, bone, or even acoustic shadow. Closer investigation of the network found that it would analyze horizontal edges in the ultrasound with greater priority than vertical edges to identify vessels with an average accuracy of $94.5\% \pm 2.9\%$.³⁶

In addition to specific structure detection in ultrasound images, researchers have also used neural networks to assist in the identification of vertebral level and other anatomical landmarks for epidural placement. Pesteie *et al.*³⁵ used convolutional neural networks to automate identification of the anterior base of the vertebral lamina,³⁵ whereas Hetherington *et al.*⁸⁰ used convolutional neural networks to automatically identify the sacrum and the L1–L5 vertebrae and vertebral spaces from ultrasound images in real time with up to 95% accuracy.⁸⁰

Pain Management. Nine ($n = 9$) papers were identified that related to the field of pain management, ranging from prediction of opioid dosing to identification of patients who may benefit from preoperative consultation with a hospital's acute pain service.^{81,82} Brown *et al.*⁸³ used machine learning to analyze differences in functional magnetic resonance imaging data collected from human volunteers who were exposed to painful and nonpainful thermal stimuli, demonstrating that machine learning analysis of whole brain scans could more accurately identify pain than analysis of individual brain regions traditionally associated with nociception.⁸³

Identification of pain was not limited to imaging techniques, Ben-Israel *et al.*⁸⁴ developed a nociception level index that was based on machine learning analysis of photoplethysmograms and skin conductance waveforms recorded from 25 patients undergoing elective surgery. However, the nociception level index was based on a target described as the combined index of stimulus and analgesia that was defined and validated within the same study, including an arbitrary ranking of intraoperative noxious stimuli.⁸⁴ In an attempt to identify more objective biomarkers for pain, Gram *et al.*⁸⁵ used machine learning to analyze electroencephalography signals from 81 patients in an attempt to predict patients who would respond to opioid therapy for acute pain; the results demonstrated only 65% accuracy of preoperative electroencephalography assessment of patients who would respond to postoperative opioid therapy.⁸⁵ Furthermore, Olesen *et al.*⁸¹ conducted a big data study searching for single nucleotide polymorphisms in 1,237 cancer patients that could predict opioid dosing for these patients; however, their study did not discover any single nucleotide polymorphism associations with opioid dosing in this population.⁸¹

Operating Room Logistics. Three ($n = 3$) papers described the use of artificial intelligence to analyze factors relating

to operating room logistics, such as scheduling of operating room time or tracking movements and actions of anesthesiologists. Combes *et al.*⁸⁶ used a hospital database containing extensive information on staffing, operating room use per procedure and staff, and post anesthesia care unit use with the electronic health record to train a neural network to predict the duration of an operation based on the team, type of operation and a patient's relevant medical history; however, prediction accuracy of their models never exceeded 60%.⁸⁶ In a different example, fuzzy logic and neural networks were used to optimize bed use for patients undergoing ophthalmologic surgery by modeling the type of case, modeling surgeon experience, staff experience, type of anesthesia and the experience of the anesthesiologist, patient factors, and comorbidities with error rates ranging from 14% to 19% depending on the type of case.⁸⁷ In a different application of machine learning, Houliston *et al.*⁸⁸ analyzed radio frequency identification tags to determine the location, orientation, and stance of anesthesiologists in the operating room. Though their analyses were limited to simulated operating rooms with mannequins, the authors proposed the use of similar tracking applications with real patients to better understand potential impacts on patient safety based on the interaction of anesthesiologists with the various equipment in the room.⁸⁸

Discussion

This scoping review identified six main clinical applications of artificial intelligence research in anesthesiology: (1) depth of anesthesia monitoring, (2) control of anesthesia, (3) event and risk prediction, (4) ultrasound guidance, (5) pain management, and (6) operating room logistics. From these applications, a summary of the most commonly encountered artificial intelligence sub-fields (*e.g.*, machine learning, computer vision, natural language processing) and methods (classical machine learning, neural networks, fuzzy logic) was presented. Most applications of artificial intelligence for anesthesiology are still in research and development; thus, the current focus of artificial intelligence within anesthesiology is not on replacing clinician judgment or skills but on investigating ways to augmenting them.

Some of the reviewed studies demonstrated improvements over existing methods and technologies in measuring clinical endpoints such as depth of anesthesia, whereas others showed no difference or even worse performance of artificial intelligence compared with currently available and widely used techniques. This dichotomy in the performance of artificial intelligence is not surprising. Although artificial intelligence as a field has been in existence for over 50 years, the recent resurgence of and increased attention on artificial intelligence as a potential enabling technology across industries has only occurred in the last 10 to 15 yr or so. The rapid advances that have been occurring in deep learning and reinforcement learning over this time frame have been attributed to a “big bang” of three factors: (1) the

availability of large datasets, (2) the advancement of hardware to perform large, parallel processing tasks (e.g., use of graphical processing units for machine learning), and (3) a new wave of development for artificial intelligence architectures and algorithms. As a result, the lay press has written of examples of impressive feats of artificial intelligence in fields such as autonomous driving, board games (e.g., chess, Go), and complex strategy-based computer games (e.g., Starcraft). However, these applications have the advantage of being constrained environments with very specifically defined rules (as in gaming) or have had catastrophic failures that have gone along with their impressive successes (as exemplified by fatalities involving autonomous vehicles).

Deep Blue's victory over chess grandmaster Garry Kasparov in 1998 stands as a famous example of a machine outperforming a human at a complex task, and even non-artificial intelligence-based computer programs have continued to demonstrate their superiority over humans in playing chess. These early programs were developed through simple optimization and search techniques or classical machine learning approaches that required input from expert players to help the machine develop effective strategies to select the best moves in a given game scenario. More recently, AlphaZero is an example of artificial intelligence that uses deep neural networks and reinforcement learning to achieve its success as it learned from each play of the game to improve its strategy. Provided just the rules of chess, Go, and shogi, and a few hours of self-play to learn strategy, AlphaZero played at a superhuman level, beating the world experts and the other top computer programs in each game.⁸⁹

We do not anticipate that artificial intelligence will master all aspects clinical anesthesiology in a matter of hours (or even years) as it has done with board games. Although games such as Go and shogi are complex, they are defined by well-understood rules that can be navigated efficiently by artificial intelligence algorithms even when the possible set of moves is incredibly large (e.g., the estimated number of possible Go games is a googolplex or $10^{(10^{100})}$).⁹⁰ Clinical medicine, on the other hand, has significant uncertainty; for much of the clinical data that we interpret during daily practice serve as surrogates for an underlying physiologic or pathophysiologic process rather than a direct measure of the process itself. Reinforcement learning and other approaches in artificial intelligence are thus left to tackle much more specific problems where the rules of clinical medicine are better understood, resulting in narrow artificial intelligence that seems constrained to accomplish one task at a time. However, feats such as artificial intelligence's conquest of complex games and recent advances in artificial intelligence developed from large, well-curated clinical datasets⁷⁰ raise hope that as we build better, more complete datasets, we could see more sophisticated and clinically meaningful applications of artificial intelligence in medicine.

As described in the Results section, one strength of machine learning is its ability to learn from data; and this learning can happen continuously as more data becomes available. This provides a distinct advantage over static algorithmic approaches (e.g., many of the existing risk prediction models that have relied on a regression model run once) that analyze data once to determine how variables might predict outcomes or other clinical factors. A learning algorithm can be used statically or continuously as data are updated, and the choice of the approach often depends on its intended use. As of April 2019, the Food and Drug Administration had approved only static or locked algorithms for medical use (e.g., detection of retinal pathology) as these algorithms can be expected to return predictable results for a given set of inputs. However, the Food and Drug Administration is considering policies for the introduction of adaptive or continuously learning algorithms based on pre-market review and clear (as yet to be determined) guidelines around data transparency and real-world performance monitoring of these algorithms.⁹¹ It is unclear which field of medicine will see the first adaptive learning algorithm approved for clinical use, but a diagnostic rather than therapeutic application may be a more likely candidate.

This review was conducted as a focused scoping review intended to assess the breadth of artificial intelligence research that has been conducted in anesthesiology rather than a systematic review designed to answer specific questions on the effectiveness or utility of artificial intelligence-based technologies in clinical practice. Although additional papers that may be relevant to artificial intelligence and anesthesiology may not have been captured with our search parameters, the breadth demonstrated in this review led us to identify areas where artificial intelligence will have implications for practicing clinicians, how clinicians can influence the future of artificial intelligence adoption, and the limitations of artificial intelligence.

Implications of Artificial Intelligence for Practicing Clinicians

The practice of modern anesthesiology requires the anesthesiologist to gather, analyze, and interpret multiple data streams for each patient. As the healthcare system has increasingly moved from analog to digital data, practicing clinicians have been asked to rely on ever-expanding, data-intensive workflows to accomplish their daily tasks. The electronic health record and the anesthesia information management system are just two of the interfaces presented to clinicians.⁹² Fortunately, information management systems that allow for automated extraction of clinical variables (e.g., vital signs, drug delivery timestamps, etc.) have eased the burden of documentation on the anesthesiologist. At the same time, the clinician must now consider how best to interpret the increasing amount of available data for the delivery of anesthetic and critical care. The application of artificial intelligence technologies should now emphasize

aiding the clinician in maximizing the clinical utility of the data that is now captured electronically.

Intraoperative and ICU monitoring of patients under anesthesia has relied on the experience of anesthesiologists to titrate anesthetics, neuromuscular blockade, and cardiovascular medications to safely maintain sedation and physiologic support. As medical technology has advanced, anesthesiologists are now expected to weigh and consider multiple sources of data to safely manage a patient's anesthesia. Although commercially available devices such as the BIS and SEDLINE (Masimo, USA) monitors have offered the promise of simplifying the assessment of hypnosis, these devices have limited algorithms that can be unreliable at extremes of age,⁹³ are dependent on the type of anesthetic agent(s) used,⁹⁴ and are susceptible to interference (from motion, electrocautery, *etc.*). On the other hand, analysis of electroencephalography data can be complex and time-consuming for anesthesiologists who are having to analyze multiple data streams.

Shortcomings and barriers to electroencephalography usage during anesthesia could be mitigated by artificial intelligence as it excels in analyzing complex, large data sets, and our review has revealed multiple ongoing efforts to identify prediction features for depth of anesthesia or consciousness states from electroencephalography signals and physiologic parameters.⁹⁵ Furthermore, because intraoperative awareness is a relatively rare event, big data approaches could be leveraged to overcome issues of data scarcity surrounding awareness. This further highlights the importance of building and curating robust datasets that can provide an appropriate source of training material for the development of future artificial intelligence systems.

Examples such as the Philips eICU database (The Netherlands) and the Multiparameter Intelligent Monitoring in Intensive Care database demonstrate how current monitors and electronic health records or information management systems can be leveraged to help automate the data collection process during episodes of care,⁹⁶ and datasets that combine traditional healthcare data (*e.g.*, claims data, vital signs, laboratory values) with specific data such as images or video could further unlock the potential of artificial intelligence algorithms that excel at analyzing and integrating vast amounts of data.⁶ With enough data, tasks such as real-time event prediction, automated adjustment of target-controlled infusions, and computer-assisted or even robotically autonomous ultrasound-guided procedures could become real possibilities. However, accurate assessment of phenomena such as intraoperative awareness will be reliant on agreed upon standards of measurement as well as good annotation of data.

Good annotation of data is critical for the success of artificial intelligence in medicine. As artificial intelligence research attempts to approximate human performance in diagnostics and therapeutics or surpass human abilities in prediction, we must remember that assessment of the

accuracy of artificial intelligence's decisions and predictions are based on accepted standards to which we compare the artificial intelligence. Some accepted standards are objective and immutable: prediction of mortality can be validated against data that has recorded patient deaths. However, other accepted standards are subject to interpretation. In the studies we reviewed on ultrasound guidance, training of the selected machine learning method for identification of structures, landmarks, *etc.*, was dependent on human labeling of the target in the training set (*i.e.*, supervised learning). Thus, assessment of the accuracy of the artificial intelligence method was also based on comparison of the machine label to that of the human, underscoring the importance of having reliable, consistent human-generated labels.

Identification of structures in ultrasounds can vary between novice and experienced sonographers. Thus, for a machine learning study or device development, it would be important to ensure that experienced sonographers were labeling training data and were the standard against which we compare a machine's performance. Similarly, as we demonstrated in our review, much of the work on monitoring depth of anesthesia in the 1990s and early 2000s was focused on the use of BIS as the target to assess awake *versus* anesthetized states in patients. BIS provided a convenient accepted standard target for artificial intelligence as a single numerical value could be used to approximate an appropriate depth of anesthesia; however, as recent clinical literature has called into question the utility of BIS in measuring depth of anesthesia, clinicians and researchers will have to collaborate to identify safe, valid target markers of hypnosis on which to train artificial intelligences.^{97,98} The continued development of technologies that use artificial intelligence, such as control systems for the delivery of medications such as anesthetics and paralytics, will be dependent on the identification of—and agreement on—accepted standard targets on which an artificial intelligence can tune its performance in achieving a given task.

Big data approaches to personalized medicine could play a large role in the delivery of anesthetic care in the future. Though a search of single nucleotide polymorphisms associated with opioid dosing was not fruitful in a study on pain management,⁸¹ artificial intelligence has been used to great effect in other fields to identify and further investigate potential genetic markers of response to specific therapies.¹² The scope of this review may not have captured current or ongoing work on genome wide association studies that could uncover markers for optimal drug selection, dosing, or adverse reactions.

At this point in time, it is hard to predict the full potential of artificial intelligence applications as we continue to make significant strides in hardware and algorithm design as well as database creation, curation, and management—advances that will undoubtedly catalyze even further advances in artificial intelligence. Before pulse oximetry, trainees were taught to recognize cyanosis and other signs of hypoxia;

before automated sphygmomanometry, adept tactile estimation of blood pressure from palpation of the pulse was a skill sought after by clinicians. Innovation led to devices that leveled the playing field for clinicians to be able to provide care based on reliable clinical metrics of oxygen saturation and blood pressure. Currently, the greatest near-term potential for artificial intelligence is in its ability to offer tools with which to analyze massive amounts of data and offer more digestible statistics about that data that clinicians can use to render a medical decision. Artificial intelligence could thus provide anesthesiologists at all levels of expertise with decision support—whether clinical or procedural—to enable all clinicians to provide the best possible evidence-based care to their patients.

As Gambus and Shafer⁹⁹ point out in their recent editorial on artificial intelligence, humans have the ability to extrapolate from the known to the unknown based on a more complete understanding of scientific phenomena, but artificial intelligence can only draw conclusions from the data it has seen and analyzed.⁹⁹ As more and more elements of clinical practice become digitized and accumulated into databases, we may one day see the development of artificial intelligence systems that have a more complete understanding of clinical phenomena and thus greater potential to deliver elements of anesthesia care autonomously.

Limitations and Ethical Implications of Artificial Intelligence

Artificial intelligence is not without its limitations. The hype surrounding artificial intelligence has reached a fever pitch in the lay press, and unrealistic expectations can result in eventual disillusionment with artificial intelligence if clinicians, patients, and regulators do not see the expected revolution in healthcare that is anticipated with artificial intelligence technologies.¹⁰⁰ Therefore, it is critical to understand that using artificial intelligence–based techniques will not necessarily result in classification or prediction that is superior to current methods. Artificial intelligence is a tool that must be deployed in the right situation to answer an appropriate question or solve an applicable problem.

One criticism of artificial intelligence, particularly of neural networks, is that the methods can result in black box results, where an algorithm can notify a clinician or researcher of a prediction but cannot provide further information on why such a prediction was made. Efforts in explainable artificial intelligence are being made to improve the transparency of algorithms. The aim of explainable artificial intelligence is to produce models that can more readily explain its findings (*e.g.*, by showing which features it may have relied on to generate its prediction) with the end goal of improving its level of transparency and therefore increasing human trust and understanding of its predictions. Some techniques in artificial intelligence will more readily lend themselves to explanation than others. For example, decision trees allow for great transparency as each decision

node could be reviewed and assessed while deep learning is currently assessed through inductive means. That is, each node within a deep learning model may not provide a clear explanation of why certain predictions were made; but the model can be asked to present relevant features or examples from its training data of skeletal x-rays to explain why a particular prediction was made on the bone age of a patient.¹⁰¹ In addition to concerns about transparency and trust in models, artificial intelligence can excel in demonstrating correlations or in identifying patterns, but it cannot yet determine causal relationships—at least not to a degree that would be necessary for clinical implementation. As research continues to expand in the field, clinicians will have to critically vet new research findings and product announcements, and medical education and training will necessarily have to incorporate elements of literacy in artificial intelligence concepts.

Artificial intelligence algorithms are also susceptible to bias in data. Beyond the basic research biases that clinicians are taught (*e.g.*, sampling, blinding, *etc.*), we must also consider both implicit and explicit bias in the healthcare system that can impact the large-scale data that is or will be used to train artificial intelligences. Eligibility of specific patient populations for clinical trials, implicit biases in treatment decisions in real world care, and other forms of bias can significantly affect the types of predictions that an artificial intelligence may make and influence clinical decisions.^{102,103} Char *et al.*¹⁰⁴ give the example of withdrawal of care from patients with traumatic brain injury. An artificial intelligence may analyze data from a neuro-ICU and interpret a pattern of fatality after traumatic brain injury as a foregone consequence of the injury rather than as secondary to the clinical decision to withdraw life support.¹⁰⁴ Therefore, it is imperative that practicing clinicians partner or develop a dialogue with data scientists to ensure the appropriate interpretation of data analyses.

Future Directions

As this review has demonstrated, artificial intelligence algorithms have not yet surpassed human performance; however, artificial intelligence's ability to quickly and accurately sift through large stores of data and uncover correlations and patterns that are imperceptible to human cognition will make it a valuable tool for clinicians. In pathology, artificial intelligence has been demonstrated to augment the ability of clinicians in making diagnoses, such as in the reduction of error rate in recognizing cancer-positive lymph nodes.³² This reduction error occurred not because the artificial intelligence outperformed humans—on the contrary it was worse—but because it could narrow the area of a histopathologic slide that a human pathologist had to review, allowing more attention to be placed on a smaller area. Similarly, we anticipate that artificial intelligence technology that can contribute to monitoring depth of anesthesia, maintaining drug infusions, or predicting hypotension in an

operation will allow practicing anesthesiologists to be more effective and efficient in the care they provide.

Medicine and the practice of anesthesiology is still, at its core, a uniquely human endeavor as both science and art. Although algorithms may one day exceed human capabilities in integrating complex, gigantic, structured datasets, much of the data that clinicians gather from patients comes from the clinician–patient relationship that is established when patients bestow trust on their doctor. Although anesthesiologists can develop the knowledge and training to trust artificial intelligence models, it remains to be seen to what extent patients will be willing to trust algorithms and how patients wish to have results from algorithms communicated to them. Therefore, qualitative research will be needed to better understand the ethical, cultural, and societal implications of integrating artificial intelligence into clinical workflows.

Anesthesiologists should continue to partner with data scientists and engineers to provide their valuable clinical insight into the development of artificial intelligence to ensure that the technology will be clinically applicable, that the data used to train algorithms are valid and representative of a wide population of patients, and that interpretations of that data are clinically meaningful.¹⁰⁵ Integration and analysis of data by models does not mean the models understand the implications of that data for specific patients. Thus, anesthesiologists should partner with other providers (*e.g.*, surgeons, interventionalists, intensivists, nurses) and patients to help develop the strategy for the optimal use of artificial intelligence. Anesthesiology as a field has been a leader in the implementation and achievement of patient safety initiatives, and artificial intelligence can serve as a new tool to continue innovations in the delivery of safe anesthesia care.

Conclusion

The field of anesthesiology has a long history of engaging in research that incorporates aspects of artificial intelligence. Artificial intelligence has the potential to impact the clinical practice of anesthesia in aspects ranging from perioperative support to critical care delivery to outpatient pain management. As research efforts advance and technology development intensifies, it will be crucial for practicing clinicians to provide practice-based insights to assist in the clinical translation of artificial intelligence.

Acknowledgments

The authors thank Elizabeth Y. Zhou, M.D. (Department of Anesthesiology and Critical Care, Perelman School of Medicine, Philadelphia, Pennsylvania), for her insights on clinical care and data streams in cardiac anesthesia and Caitlin Stafford, B.S., C.C.R.P. (Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts) for her insights on implementation and management of large clinical databases.

Research Support

Support for this study was provided solely from institutional and/or departmental sources.

Competing Interests

Dr. Hashimoto is a consultant for Verily Life Sciences (Mountain View, California), the Johnson & Johnson Institute (Raynham, Massachusetts), and Gerson Lehrman Group (New York, New York). He serves on the clinical advisory board of Worrell, Inc. (Minneapolis, Minnesota). Dr. Meireles is a consultant for Olympus (Tokyo, Japan) and Medtronic (Dublin, Ireland). Dr. Rosman is a research scientist at Toyota Research Institute (Cambridge, Massachusetts). Drs. Hashimoto, Meireles, and Rosman are co-inventors on a patent application submitted in relation to computer vision technology for assessment of intraoperative surgical performance. The other authors declare no competing interests.

Correspondence

Address correspondence to Dr. Hashimoto: Surgical Artificial Intelligence and Innovation Laboratory, Massachusetts General Hospital, 15 Parkman Street, WAC 339, Boston, Massachusetts 02139. dahashimoto@mgh.harvard.edu. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

References

1. Bellman R: An introduction to artificial intelligence: Can computers think? San Francisco, Boyd & Fraser Pub Co, 1978
2. Buchanan BG: A (very) brief history of artificial intelligence. *Ai Magazine* 2005; 26: 53
3. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, Brink J: Artificial intelligence and machine learning in radiology: Opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol* 2018; 15(3 Pt B):504–8
4. Salto-Tellez M, Maxwell P, Hamilton P: Artificial intelligence—the third revolution in pathology. *Histopathology* 2019; 74:372–6
5. Deo RC: Machine learning in medicine. *Circulation* 2015; 132:1920–30
6. Hashimoto DA, Rosman G, Rus D, Meireles OR: Artificial intelligence in surgery: Promises and perils. *Ann Surg* 2018; 268:70–6
7. FDA: FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. Edited by FDA. US Food and Drug Administration, 2018

8. Russell S, Norvig P: Artificial Intelligence: A Modern Approach, 3rd Edition. Upper Saddle River, New Jersey, Prentice Hall, 2009
9. Kendale S, Kulkarni P, Rosenberg AD, Wang J: Supervised machine-learning predictive analytics for prediction of postinduction hypotension. *ANESTHESIOLOGY* 2018; 129:675–88
10. Wanderer JP, Rathmell JP: Machine learning for anesthesiologists: A primer. *ANESTHESIOLOGY* 2018; 129:A29
11. Bisgin H, Liu Z, Fang H, Xu X, Tong W: Mining FDA drug labels using an unsupervised learning technique—topic modeling. *BMC Bioinformatics* 2011; 12 Suppl 10:S11
12. Hakonarson H, Bjornsdottir US, Halapi E, Bradfield J, Zink F, Mouy M, Helgadottir H, Gudmundsdottir AS, Andrason H, Adalsteinsdottir AE, Kristjansson K, Birkisson I, Arnason T, Andresdottir M, Gislason D, Gislason T, Gulcher JR, Stefansson K: Profiling of genes expressed in peripheral blood mononuclear cells predicts glucocorticoid sensitivity in asthma patients. *Proc Natl Acad Sci U S A* 2005; 102:14789–94
13. Sutton RS, Barto AG: Reinforcement Learning: An Introduction, MIT press Cambridge, 1998
14. Padmanabhan R, Meskin N, Haddad WM: Closed-loop control of anesthesia and mean arterial pressure using reinforcement learning. *Biomed Signal Process Control* 2015; 22: 54–64
15. Ng A: Supervised Learning, CS229 Lecture Notes. Edited by Ng A, Stanford University, 2018
16. Zadeh LA: Fuzzy sets. *Information and Control* 1965; 8: 338–353
17. Baig MM, Gholamhosseini H, Kouzani A, Harrison MJ: Anaesthesia monitoring using fuzzy logic. *J Clin Monit Comput* 2011; 25:339–47
18. Hu YJ, Ku TH, Jan RH, Wang K, Tseng YC, Yang SF: Decision tree-based learning to predict patient controlled analgesia consumption and readjustment. *BMC Med Inform Decis Mak* 2012; 12:131
19. Hastie T, Tibshirani R, Friedman J: Overview of Supervised Learning, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, Springer, 2016, pp 9–42
20. Hastie T, Tibshirani R, Friedman J: Support Vector Machines and Flexible Discriminants, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, Springer, 2016, pp 417–58
21. McCulloch WS, Pitts W: A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 1943; 5: 115–33
22. LeCun Y, Bengio Y, Hinton G: Deep learning. *Nature* 2015; 521:436–44
23. Bland JM, Altman DG: Bayesians and frequentists. *BMJ* 1998; 317:1151–60
24. Ghahramani Z: Probabilistic machine learning and artificial intelligence. *Nature* 2015; 521:452–9
25. Kirkos E, Spathis C, Manolopoulos Y: Data mining techniques for the detection of fraudulent financial statements. *Expert Syst Appl* 2007; 32: 995–1003
26. van den Berg JP, Eleveld DJ, De Smet T, van den Heerik AVM, van Amsterdam K, Lichtenbelt BJ, Scheeren TWL, Absalom AR, Struys MMRF: Influence of Bayesian optimization on the performance of propofol target-controlled infusion. *Br J Anaesth* 2017; 119:918–27
27. Kukacka M: Bayesian Methods in Artificial Intelligence, WDS'10 Proceedings of Contributed Papers, 2010, pp 25–30
28. Nadkarni PM, Ohno-Machado L, Chapman WW: Natural language processing: An introduction. *J Am Med Inform Assoc* 2011; 18:544–51
29. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, Dittus RS, Rosen AK, Elkin PL, Brown SH, Speroff T: Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 2011; 306:848–55
30. Cohen KB, Glass B, Greiner HM, Holland-Bouley K, Standridge S, Arya R, Faist R, Morita D, Mangano F, Connolly B, Glauser T, Pestian J: Methodological issues in predicting pediatric epilepsy surgery candidates through natural language processing and machine learning. *Biomed Inform Insights* 2016; 8:11–8
31. Friedman C, Shagina L, Lussier Y, Hripcsak G: Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc* 2004; 11:392–402
32. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, van der Laak JAWM, Hermsen M, Manson QF, Balkenhol M, Geessink O, Stathonikos N, van Dijk MC, Bult P, Beca F, Beck AH, Wang D, Khosla A, Gargeya R, Irshad H, Zhong A, Dou Q, Li Q, Chen H, Lin HJ, Heng PA, Haß C, Bruni E, Wong Q, Halici U, Öner MÜ, Cetin-Atalay R, Berseth M, Khvatkov V, Vylegzhanin A, Kraus O, Shaban M, Rajpoot N, Awan R, Sirinukunwattana K, Qaiser T, Tsang YW, Tellez D, Annuschein J, Hufnagl P, Valkonen M, Kartasalo K, Latonen L, Ruusuuvuori P, Liimatainen K, Albarqouni S, Mungal B, George A, Demirci S, Navab N, Watanabe S, Seno S, Takenaka Y, Matsuda H, Ahmady Phoulady H, Kovalev V, Kalinovskiy A, Liauchuk V, Bueno G, Fernandez-Carrobles MM, Serrano I, Deniz O, Racoceanu D, Venâncio R; the CAMELYON16 Consortium: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; 318:2199–210
33. Tajmir SH, Lee H, Shailam R, Gale HI, Nguyen JC, Westra SJ, Lim R, Yune S, Gee MS, Do S: Artificial intelligence-assisted interpretation of bone age radiographs improves accuracy and decreases variability. *Skeletal Radiol* 2019; 48:275–83

34. Volkov M, Hashimoto DA, Rosman G, Meireles OR, Rus D: Machine learning and coresets for automated real-time video segmentation of laparoscopic and robot-assisted surgery, IEEE International Conference on Robotics and Automation (ICRA), 2017, pp 754–9
35. Pesteie M, Lessoway V, Abolmaesumi P, Rohling RN: Automatic localization of the needle target for ultrasound-guided epidural injections. *IEEE Trans Med Imaging* 2018; 37:81–92
36. Smistad E, Lovstakken L, Carneiro G, Mateus D, Peter L, Bradley A, Tavares J, Belagiannis V, Papa JP, Nascimento JC, Loog M, Lu Z, Cardoso JS, Cornebise J: Vessel Detection in Ultrasound Images Using Deep Convolutional Neural Networks, *Med Image Comput Comput Assist Inter*, Springer, 2016, pp 30–8
37. Fritz BA, Maybrier HR, Avidan MS: Intraoperative electroencephalogram suppression at lower volatile anaesthetic concentrations predicts postoperative delirium occurring in the intensive care unit. *Br J Anaesth* 2018; 121:241–8
38. Kertai MD, Pal N, Palanca BJ, Lin N, Searleman SA, Zhang L, Burnside BA, Finkel KJ, Avidan MS; B-Unaware Study Group: Association of perioperative risk factors and cumulative duration of low bispectral index with intermediate-term mortality after cardiac surgery in the B-Unaware Trial. *ANESTHESIOLOGY* 2010; 112:1116–27
39. Sessler DI, Sigl JC, Kelley SD, Chamoun NG, Manberg PJ, Saager L, Kurz A, Greenwald S: Hospital stay and mortality are increased in patients having a “triple low” of low blood pressure, low bispectral index, and low minimum alveolar concentration of volatile anesthesia. *ANESTHESIOLOGY* 2012; 116:1195–203
40. Veselis RA, Reinsel R, Sommer S, Carlon G: Use of neural network analysis to classify electroencephalographic patterns against depth of midazolam sedation in intensive care unit patients. *J Clin Monit* 1991; 7:259–67
41. Veselis RA, Reinsel R, Wronski M: Analytical methods to differentiate similar electroencephalographic spectra: neural network and discriminant analysis. *J Clin Monit* 1993; 9:257–67
42. Ortolani O, Conti A, Di Filippo A, Adembri C, Moraldi E, Evangelisti A, Maggini M, Roberts SJ: EEG signal processing in anaesthesia: Use of a neural network technique for monitoring depth of anaesthesia. *Br J Anaesth* 2002; 88:644–8
43. Benzy VK, Jasmin EA, Koshy RC, Amal F: Wavelet Entropy based classification of depth of anesthesia. 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT) 2016: 521–4
44. Mirsadeghi M, Behnam H, Shalbfaf R, Jelveh Moghadam H: Characterizing awake and anesthetized states using a dimensionality reduction method. *J Med Syst* 2016; 40:13
45. Shalbfaf A, Saffar M, Sleight JW, Shalbfaf R: Monitoring the depth of anesthesia using a new adaptive neuro-fuzzy system. *IEEE J Biomed Health Inform* 2018; 22:671–7
46. Nagaraj SB, Biswal S, Boyle EJ, Zhou DW, McClain LM, Bajwa EK, Quraishi SA, Akeju O, Barbieri R, Purdon PL, Westover MB: Patient-specific classification of ICU sedation levels from heart rate variability. *Crit Care Med* 2017; 45:e683–90
47. Ranta SO, Hynynen M, Räsänen J: Application of artificial neural networks as an indicator of awareness with recall during general anaesthesia. *J Clin Monit Comput* 2002; 17:53–60
48. Dumont GA, Ansermino JM: Closed-loop control of anesthesia: A primer for anesthesiologists. *Anesth Analg* 2013; 117:1130–8
49. Tsutsui T, Arita S: Fuzzy-logic control of blood pressure through enflurane anesthesia. *J Clin Monit* 1994; 10:110–7
50. Zbinden AM, Feigenwinter P, Petersen-Felix S, Hacidalilzade S: Arterial pressure control with isoflurane using fuzzy logic. *Br J Anaesth* 1995; 74:66–72
51. Absalom AR, Sutcliffe N, Kenny GN: Closed-loop control of anesthesia using Bispectral index: Performance assessment in patients undergoing major orthopedic surgery under combined general and regional anesthesia. *ANESTHESIOLOGY* 2002; 96:67–73
52. Shieh JS, Kao MH, Liu CC: Genetic fuzzy modeling and control of bispectral index (BIS) for general intravenous anaesthesia. *Med Eng Phys* 2006; 28:134–48
53. Lowery C, Faisal AA: Towards efficient, personalized anesthesia using continuous reinforcement learning for propofol infusion control. 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER) 2013: 1414–7
54. Zaouter C, Hemmerling TM, Lanchon R, Valoti E, Remy A, Leuillet S, Ouattara A: The feasibility of a completely automated total IV anesthesia drug delivery system for cardiac surgery. *Anesth Analg* 2016; 123:885–93
55. Lendl M, Schwarz UH, Romeiser HJ, Unbehauen R, Georgieff M, Geldner GF: Nonlinear model-based predictive control of non-depolarizing muscle relaxants using neural networks. *J Clin Monit Comput* 1999; 15:271–8
56. Shieh JS, Fan SZ, Chang LW, Liu CC: Hierarchical rule-based monitoring and fuzzy logic control for neuromuscular block. *J Clin Monit Comput* 2000; 16:583–92
57. Motamed C, Devys JM, Debaene B, Billard V: Influence of real-time Bayesian forecasting of pharmacokinetic parameters on the precision of a rocuronium target-controlled infusion. *Eur J Clin Pharmacol* 2012; 68:1025–31

58. Martinoni EP, Pfister ChA, Stadler KS, Schumacher PM, Leibundgut D, Bouillon T, Böhlen T, Zbinden AM: Model-based control of mechanical ventilation: Design and clinical validation. *Br J Anaesth* 2004; 92:800–7
59. Schäublin J, Derighetti M, Feigenwinter P, Petersen-Felix S, Zbinden AM: Fuzzy logic control of mechanical ventilation during anaesthesia. *Br J Anaesth* 1996; 77:636–41
60. Schädler D, Mersmann S, Frerichs I, Elke G, Semmel-Griebeler T, Noll O, Pulletz S, Zick G, David M, Heinrichs W, Scholz J, Weiler N: A knowledge- and model-based system for automated weaning from mechanical ventilation: Technical description and first clinical application. *J Clin Monit Comput* 2014; 28:487–98
61. Lin CS, Li YC, Mok MS, Wu CC, Chiu HW, Lin YH: Neural network modeling to predict the hypnotic effect of propofol bolus induction. *Proc AMIA Symp* 2002: 450–3
62. Nunes CS, Mendonca TF, Amorim P, Ferreira DA, Antunes L: Comparison of neural networks, fuzzy and stochastic prediction models for return of consciousness after general anesthesia. *Proceedings of the 44th IEEE Conference on Decision and Control* 2005: 4827–32
63. Santanen OA, Svartling N, Haasio J, Paloheimo MP: Neural nets and prediction of the recovery rate from neuromuscular block. *Eur J Anaesthesiol* 2003; 20:87–92
64. Lin CS, Chang CC, Chiu JS, Lee YW, Lin JA, Mok MS, Chiu HW, Li YC: Application of an artificial neural network to predict postinduction hypotension during general anesthesia. *Med Decis Making* 2011; 31:308–14
65. Lin CS, Chiu JS, Hsieh MH, Mok MS, Li YC, Chiu HW: Predicting hypotensive episodes during spinal anesthesia with the application of artificial neural networks. *Comput Methods Programs Biomed* 2008; 92:193–7
66. Zhang L, Fabbri D, Lasko TA, Ehrenfeld JM, Wanderer JP: A system for automated determination of perioperative patient acuity. *J Med Syst* 2018; 42:123
67. Moustafa MA, El-Metainy S, Mahar K, Mahmoud Abdel-magied E: Defining difficult laryngoscopy findings by using multiple parameters: A machine learning approach. *Egypt J Anaesth* 2017; 33: 153–8
68. Berkenstadt H, Ben-Menachem E, Herman A, Dach R: An evaluation of the Integrated Pulmonary Index (IPI) for the detection of respiratory events in sedated patients undergoing colonoscopy. *J Clin Monit Comput* 2012; 26:177–81
69. Hancerliogullari G, Hancerliogullari KO, Koksalmis E: The use of multi-criteria decision making models in evaluating anesthesia method options in circumcision surgery. *BMC Med Inform Decis Mak* 2017; 17:14
70. Hatib F, Jian Z, Buddi S, Lee C, Settels J, Sibert K, Rinehart J, Cannesson M: Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *ANESTHESIOLOGY* 2018; 129:663–74
71. Gao L, Smielewski P, Czosnyka M, Ercole A: Cerebrovascular signal complexity six hours after intensive care unit admission correlates with outcome after severe traumatic brain injury. *J Neurotrauma* 2016; 33:2011–8
72. Gao L, Smielewski P, Czosnyka M, Ercole A: Early asymmetric cardio-cerebral causality and outcome after severe traumatic brain injury. *J Neurotrauma* 2017; 34:2743–52
73. Gottschalk A, Hyzer MC, Geer RT: A comparison of human and machine-based predictions of successful weaning from mechanical ventilation. *Med Decis Making* 2000; 20:160–9
74. Bonds BW, Yang S, Hu PF, Kalpakis K, Stansbury LG, Scalea TM, Stein DM: Predicting secondary insults after severe traumatic brain injury. *J Trauma Acute Care Surg* 2015; 79:85–90; discussion 90
75. Jalali A, Bender D, Rehman M, Nadkanri V, Nataraj C: Advanced analytics for outcome prediction in intensive care units. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2016: 2520–4
76. Clermont G, Angus DC, DiRusso SM, Griffin M, Linde-Zwirble WT: Predicting hospital mortality for patients in the intensive care unit: A comparison of artificial neural networks with logistic regression models. *Crit Care Med* 2001; 29:291–6
77. Desautels T, Das R, Calvert J, Trivedi M, Summers C, Wales DJ, Ercole A: Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: A cross-sectional machine learning approach. *BMJ Open* 2017; 7:e017199
78. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, Shimabukuro D, Chettipally U, Feldman MD, Barton C, Wales DJ, Das R: Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach. *JMIR Med Inform* 2016; 4:e28
79. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R: Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: A randomised clinical trial. *BMJ Open Respir Res* 2017; 4:e000234
80. Hetherington J, Lessoway V, Gunka V, Abolmaesumi P, Rohling R: SLIDE: Automatic spine level identification system using a deep convolutional neural network. *Int J Comput Assist Radiol Surg* 2017; 12:1189–98
81. Olesen AE, Grønlund D, Gram M, Skorpen F, Drewes AM, Klepstad P: Prediction of opioid dose in cancer

- pain patients using genetic profiling: Not yet an option with support vector machine learning. *BMC Res Notes* 2018; 11:78
82. Tighe PJ, Lucas SD, Edwards DA, Boezaart AP, Aytug H, Bihorac A: Use of machine-learning classifiers to predict requests for preoperative acute pain service consultation. *Pain Med* 2012; 13:1347–57
 83. Gonzalez-Cava JM, Arnay R, Perez JAM, Leon A, Martin M, Jove-Perez E, Calvo-Rolle JL, Casteleiro-Roca JL, Juez FJD, Garcia HP, AlfonsoCendon J, Gonzalez LS, Quintian H, Corchado E: A Machine learning based system for analgesic drug delivery. International joint Conference SOCO'17-CISIS'17-ICEUTE'17. 2017; 649: 461–70
 84. Ben-Israel N, Kliger M, Zuckerman G, Katz Y, Edry R: Monitoring the nociception level: A multi-parameter approach. *J Clin Monit Comput* 2013; 27:659–68
 85. Gram M, Erlenwein J, Petzke F, Falla D, Przemek M, Emons MI, Reuster M, Olesen SS, Drewes AM: Prediction of postoperative opioid analgesia using clinical-experimental parameters and electroencephalography. *Eur J Pain* 2017; 21:264–77
 86. Combes C, Meskens N, Rivat C, Vandamme JP: Using a KDD process to forecast the duration of surgery. *Int J Prod Econ* 2008; 112: 279–93
 87. Devi SP, Rao KS, Sangeetha SS: Prediction of surgery times and scheduling of operation theaters in ophthalmology department. *J Med Syst* 2012; 36:415–30
 88. Houliston BR, Parry DT, Merry AF: TADAA: Towards automated detection of anaesthetic activity. *Methods Inf Med* 2011; 50:464–71
 89. Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, Lanctot M, Sifre L, Kumaran D, Graepel T, Lillicrap T, Simonyan K, Hassabis D: A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* 2018; 362:1140–4
 90. Walraet M, Tromp J: *A Googolplex of Go Games*. Cham, Springer International Publishing, 2016, pp 191–201
 91. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) – Discussion Paper and Request for Feedback, Food and Drug Administration, 2019, pp 1–20
 92. Liem VGB, Hoeks SE, van Lier F, de Graaff JC: What we can learn from Big Data about factors influencing perioperative outcome. *Curr Opin Anaesthesiol* 2018; 31:723–31
 93. Sciusco A, Standing JF, Sheng Y, Raimondo P, Cinnella G, Dambrosio M: Effect of age on the performance of bispectral and entropy indices during sevoflurane pediatric anesthesia: A pharmacometric study. *Paediatr Anaesth* 2017; 27:399–408
 94. Olofsen E, Sleight JW, Dahan A: Permutation entropy of the electroencephalogram: A measure of anaesthetic drug effect. *Br J Anaesth* 2008; 101:810–21
 95. Eagleman SL, Drover DR: Calculations of consciousness: electroencephalography analyses to determine anesthetic depth. *Curr Opin Anaesthesiol* 2018; 31:431–8
 96. Ghassemi M, Celi LA, Stone DJ: State of the art review: The data revolution in critical care. *Crit Care* 2015; 19:118
 97. Avidan MS, Jacobsohn E, Glick D, Burnside BA, Zhang L, Villafranca A, Karl L, Kamal S, Torres B, O'Connor M, Evers AS, Gradwohl S, Lin N, Palanca BJ, Mashour GA; BAG-RECALL Research Group: Prevention of intraoperative awareness in a high-risk surgical population. *N Engl J Med* 2011; 365:591–600
 98. Avidan MS, Zhang L, Burnside BA, Finkel KJ, Searleman AC, Selvidge JA, Saager L, Turner MS, Rao S, Bottros M, Hantler C, Jacobsohn E, Evers AS: Anesthesia awareness and the bispectral index. *N Engl J Med* 2008; 358:1097–108
 99. Gambus P, Shafer SL: Artificial intelligence for everyone. *ANESTHESIOLOGY* 2018; 128:431–3
 100. Sicular S, Brant K: *Hype Cycle for Artificial Intelligence, 2018*. Edited by Gartner I. Gartner, 2018, pp 1–73
 101. Lee H, Tajmir S, Lee J, Zissen M, Yeshiwas BA, Alkasab TK, Choy G, Do S: Fully automated deep learning system for bone age assessment. *J Digit Imaging* 2017; 30:427–41
 102. Murthy VH, Krumholz HM, Gross CP: Participation in cancer clinical trials: Race-, sex-, and age-based disparities. *JAMA* 2004; 291:2720–6
 103. Schulman KA, Berlin JA, Harless W, Kerner JF, Sistrunk S, Gersh BJ, Dubé R, Taleghani CK, Burke JE, Williams S, Eisenberg JM, Escarce JJ: The effect of race and sex on physicians' recommendations for cardiac catheterization. *N Engl J Med* 1999; 340:618–26
 104. Char DS, Shah NH, Magnus D: Implementing machine learning in health care – Addressing ETHICAL CHALLENGES. *N Engl J Med* 2018; 378:981–3
 105. Weber GM, Mandl KD, Kohane IS: Finding the missing link for big biomedical data. *JAMA* 2014; 311:2479–800