Review

Check for updates

The importance of machine learning in autonomous actions for surgical decision making

Martin Wagner[1], Sebastian Bodenstedt[2,5], Marie Daum[1], Andre Schulze[1], Rayan Younis[1], Johanna Brandenburg[1], Fiona R. Kolbinger[3], Marius Distler[3,5], Lena Maier-Hein[4], Jürgen Weitz[3,5], Beat-Peter Müller-Stich[1], Stefanie Speidel[2,5]

[1]Department for General, Visceral, and Transplantation Surgery, Heidelberg University Hospital, Heidelberg 69120 Germany.
[2]Division of Translational Surgical Oncology, National Center for Tumor Diseases (NCT/UCC), Dresden 01307, Germany.
[3]Department for Visceral, Thoracic, and Vascular Surgery at the University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden 01307, Germany.
[4]Division of Computer Assisted Medical Interventions (CAMI), German Cancer Research Center (DKFZ), Heidelberg 69120, Germany.
[5]Centre for Tactile Internet with Human-in-the-Loop (CeTI), TU Dresden, Dresden 01069, Germany.

Correspondence to: Prof. Stefanie Speidel, Division of Translational Surgical Oncology, National Center for Tumor Diseases (NCT/UCC), Fetscherstr. 74, Dresden 01307, Germany. E-mail: stefanie.speidel@nct-dresden.de

Abstract

Surgery faces a paradigm shift since it has developed rapidly in recent decades, becoming a high-tech discipline. Increasingly powerful technological developments such as modern operating rooms, featuring digital and interconnected equipment and novel imaging as well as robotic procedures, provide several data sources resulting in a huge potential to improve patient therapy and surgical outcome by means of Surgical Data Science. The emerging field of Surgical Data Science aims to improve the quality of surgery through acquisition, organization, analysis, and modeling of data, in particular using machine learning (ML). An integral part of surgical data science is to analyze the available data along the surgical treatment path and provide a context-aware autonomous action by means of ML methods. Autonomous actions related to surgical decision-making include preoperative decision support, intraoperative assistance functions, as well as robot-assisted actions. The goal is to democratize surgical skills and enhance the collaboration between surgeons and cyber-physical systems by quantifying surgical experience and making it accessible to machines, thereby improving patient therapy and outcome. The article introduces basic ML concepts as enablers for autonomous actions in surgery, highlighting examples for such

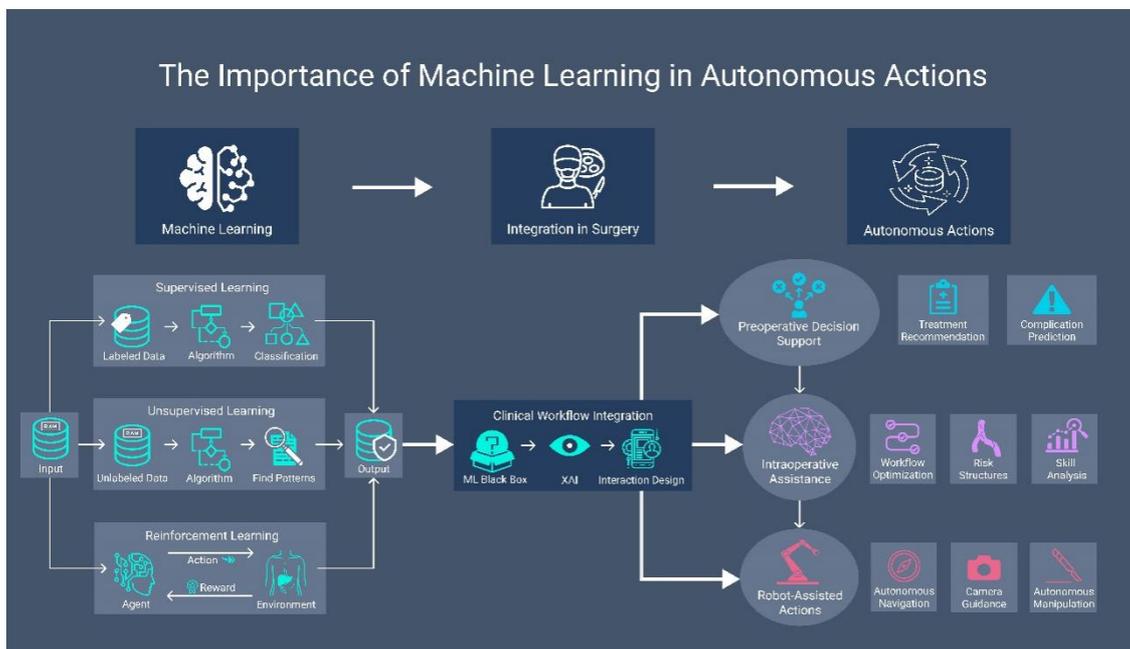actions along the surgical treatment path.

INTRODUCTION

Surgical complications are the third leading contributor to death worldwide[1]. According to current studies, around nine million serious complications occur in an estimated 300 million operations each year[2]. Surgery faces a paradigm shift since it has developed rapidly in recent decades, becoming a high-tech discipline. Increasingly powerful technological developments such as modern operating rooms (OR), featuring digital and interconnected equipment and novel imaging as well as robotic procedures, provide a huge potential to improve patient therapy and surgical outcome. Although such devices and data are available, the human ability to use these possibilities, especially in a complex and time-critical situation such as surgery, is limited and is extremely dependent on the experience of the surgical staff[3].

Machine learning (ML), in particular deep learning, has recently seen major success in the medical domain, such as prediction of kidney failure[4], detection of molecular alterations on histopathology slides[5], or skin lesion classification[6]. The field of surgery is, however, not affected in a similar fashion yet. In surgery, unlike the aforementioned examples, we have procedural multimodal data in a dynamic environment including the patient, different devices and sensors in the OR, and the OR team; of course, the high variability of the patient (e.g., anatomy, tumor localization, comorbidities, *etc.*) and surgical team (e.g., surgical procedure, surgical strategy, surgical experience, *etc.*) characteristics have to be considered as well. This poses several challenges regarding technical infrastructure for data acquisition, storage and access, data annotation and sharing, data analytics, and aspects related to clinical translation as identified by the Surgical Data Science Initiative[7]. The international Surgical Data Science Initiative (http://www.surgical-data-science.org/) was founded in 2015 with the mission to pave the way for artificial intelligence (AI) success stories in surgery by "improving the quality of interventional healthcare and its value through capturing, organization, analysis, and modeling of data"[8]. An integral part of surgical data science is to analyze the available data along the surgical treatment path and provide a context-aware autonomous action by means of ML methods. Autonomous actions related to surgical decision-making include preoperative decision support, intraoperative assistance functions, as well as robot-assisted actions. The goal is to democratize surgical skills and enhance the collaboration between surgeons and cyber-physical systems by quantifying surgical experience and making it accessible to machines, thereby improving patient therapy and outcome.

In the following, basic ML concepts will be introduced as enablers for autonomous actions in surgery, highlighting examples for such actions along the surgical treatment path. Moreover, we will discuss ongoing challenges in the field of AI-based surgery. Figure 1 provides an overview of the importance of ML in autonomous actions in surgery.

MACHINE LEARNING AS AN ENABLER FOR AUTONOMOUS ACTIONS

AI is a research area in computer science working on developing intelligent agents in the form of computer programs that aim to mimic the cognitive functions of humans, that is, by acting or by thinking like a human. ML is a sub-field of AI that focuses on imitating the human ability to learn. For this, ML relies mainly on data and using algorithms to learn from this data and draw conclusions, such as how to act in a traffic situation or distinguish healthy tissue from cancerous tissue. ML comprises different algorithms that can be used to learn from data. ML algorithms are generally divided into three different categories based on

Figure 1. Autonomous actions in surgery along the surgical treatment path based on machine learning and clinical integration as a prerequisite.

their level of supervision: unsupervised, supervised, and reinforcement learning[9].

Supervised learning

Supervised learning methods require labeled data. This means that a label has to be provided for each data point in a dataset, such as an image, time series data, or sound. Based on the labeled data, supervised learning methods will then learn a function that maps data points onto those labels. Labeled data is often acquired through manual annotation by experts, though approaches for automatically generating labels from unlabeled data exist. An example might be that a laparoscopic image showing a liver might be associated with the label "liver", while an image of a "kidney" should be associated with the label "kidney". Supervised learning methods analyze the data to determine patterns that will allow the algorithm to generalize from the given, labeled data to new data, thereby being able to assign given, unlabeled data the correct label. Supervised learning methods can deal with symbolic labels (classification), such as "liver" or "non-cancerous tissue," and with numerical sub-symbolic labels (regression), such as temperatures, size, or age. Some examples of supervised learning methods are random forests, neural networks, or support vector machines[9]. In surgery, supervised learning may be used to detect the presence of instruments or organs in laparoscopic surgery videos or predict survival after cancer surgery.

Unsupervised learning

In contrast to supervised learning methods, unsupervised learning methods do not require labeled data. Unsupervised learning methods aim to discover patterns in unlabeled data, with the goal of learning useful representations of the data. Such representations could then be used to determine irregularities in data (e.g., to discover fraud in credit card data using a clustering method such as k-means) or as a pre-processing step for supervised learning (e.g., by realigning data to make it easier separable using the Principal Component Analysis)[9]. In surgery, unsupervised learning may be used to detect deviations from a standardized treatment path that may lead to complications.

Reinforcement learning
Algorithms for supervised learning learn from annotated examples, meaning that data must be labeled in order to be accessible to the algorithms. Labeling a sufficient amount of data is not always feasible. Consider teaching an agent how to play a complex game like chess, which possesses an almost immeasurable number of possible combinations of board states (about 1040). It is infeasible to collect labels for anything but an insignificant fraction of these states, describing what might be a "good" or a "bad" move in each case. An alternative for learning to navigate such complex problems is reinforcement learning. In contrast to supervised learning, reinforcement learning does not require labeled data; instead, it depends on rewards, which occur depending on what actions (e.g., a chess move) are taken in an environment (e.g., a game board). Rewards are awarded after an action was taken, for example, a reward of "1" after the winning move in a game of chess or of "0" after the losing move. The aim of reinforcement learning is then to learn a policy that enables the agent to take actions that should maximize the cumulative reward[10,9]. In surgery, reinforcement may be used to train surgical robots how to perform context-aware autonomous actions in laparoscopic surgery.

Neural networks and deep learning
Artificial neural networks are a subset of ML inspired by biological neurons and the connections they form. Artificial neural networks generally consist of multiple layers of artificial neurons, where neurons are connected to neurons of one another. These layers can be divided into input layers, output layers, and intermediate layers (also called hidden layers), which connect the input to the output [see Figure 1]. If the input is an image, the first layer could, for example, learn to detect straight lines in the image, and the next layer might use those lines to determine if certain geometric primitives are visible. Depending on the flow of data, neural networks can be divided into feed-forward networks [see Figure 2 Left], in which the data only flows into the direction of the output, and recurrent networks [see Figure 2 Right], in which data from hidden and output layers can flow back into the network as input. Recurrent networks therefore possess an internal state, or memory, while feed-forward networks are stateless[9,11]. For analyzing data in which spatial relationships exist, such as time series data, images, videos, and volume data, a specialized layer type, the convolutional layer, was developed[12,13]. These layers incorporate spatial information on data points (e.g., which pixels in an image are neighbors), and make it possible to analyze these data points within their spatial context, providing a powerful tool for data analysis. Neural networks that include such convolutional layers are referred to as convolutional neural networks (CNNs).

Neural networks with multiple hidden layers are considered deep neural networks and form the basis of "deep learning". Increasing the depth of a neural network can learn more complex features and model more complex problems. While more depth can potentially increase the performance of neural networks, it also leads to more open parameters that must be learned, which means that more data is required to train the network.

As deep learning requires vast amounts of data for training, new approaches to utilizing unlabeled data during training are actively being developed. One approach to employing unlabeled data is so-called "self-supervised learning". This generally is a two-step approach. First, a neural network is trained to solve a secondary task on unlabeled data by leveraging labels inherent to the data. Examples for such a task could be learning to colorize grayscale laparoscopic images[14] or determining if and how far an image was rotated[15]. In a second step, these neural networks are then trained further on labeled data for the primary task, such as segmenting instruments in laparoscopic images or identifying objects in natural images. The rationale behind this approach is that by first solving the secondary task, the network has to learn to extract features that are useful in solving the primary task, thereby boosting performance.
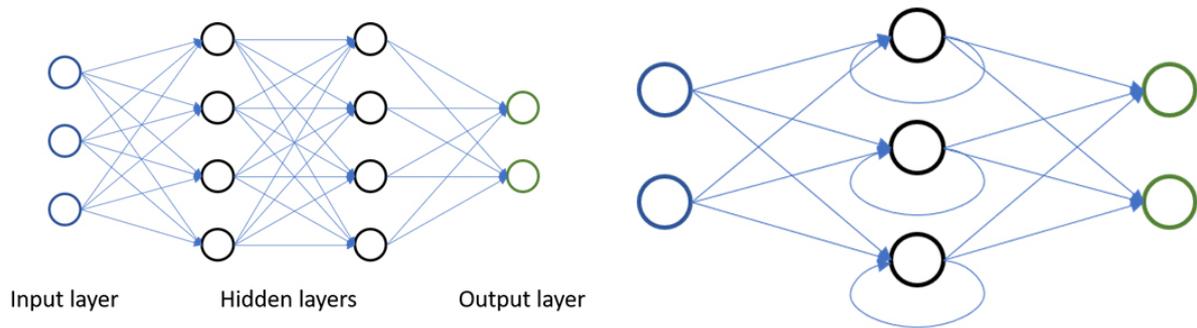
Figure 2. Left: Feed-forward neural network with two hidden layers; Right: recurrent neural network.

A different approach to coping with limited data that current research focuses on is few-shot learning. In few-shot learning, the model is tasked with learning to classify new data by only seeing a few training examples. One way of accomplishing this is via meta-learning, where a model is trained on several different tasks simultaneously to help solve unrelated tasks[16].

This ability to model complex problems makes deep neural networks extremely powerful and versatile, so they are currently state of the art for many problems in ML, such as computer vision tasks like image and video recognition, natural language processing, and reinforcement learning[9,17].

Uncertainty & explainable AI

Two of the bigger issues in applying ML algorithms, especially deep learning, to problems in critical fields, such as surgery, are determining how certain (or uncertain) the algorithm is of its prediction and determining why the algorithm arrived at a certain prediction. For example, an algorithm trained to distinguish between organs in laparoscopic surgery, such as "liver" and "kidney." While the algorithm can solve this problem with high accuracy, its behavior is unpredictable when given an image of anything else, such as images from a urological surgery where the prostate can be seen. Standard deep neural networks, for example, generally do not have an output option for "I don't know" and will therefore sort any given data into the categories that it is aware of. To help deal with such errors, methods for dealing with uncertainty in deep learning have been developed (e.g., Bayesian deep learning)[18]. When talking about uncertainty quantification, literature distinguishes between epistemic and aleatoric uncertainty. Epistemic uncertainty arises due to inappropriate or incomplete training data (i.e., ignorance of our model towards certain data points), a problem that can be solved with more training. The prior example is a case of epistemic uncertainty. Aleatoric uncertainty, on the other hand, arises from stochastic processes (e.g., measurement errors or sensor noise) and cannot be combated with more data[19]. The third type of uncertainty that is gaining increasing attention in the literature is the potential inherent ambiguity of a problem. State-of-the-art ML approaches typically provide so-called point estimates (i.e., one solution) and neglect the fact that the problem may be inherently ambiguous. For example, matching intra-operative limited-view data to preoperative data may have two different, mathematically similarly plausible solutions. For instance, such problems can be addressed with invertible neural networks[20,21].

Due to the complexity of deep learning models, their output is not easily interpretable, rendering them something akin to a black box. Recent research in the area of eXplainable AI (XAI) aims to make the predictions of ML models, especially deep learning models, and the reasoning behind those predictions more transparent. Here different approaches are being explored, for example generating text-based explanations of the results, visual explanation techniques that attempt to visualize how a model reached a

conclusion, or via explanation by example, where representative examples are used to illustrate the reasoning of a model[22].

AUTONOMOUS ACTIONS IN SURGERY

Autonomous actions related to surgical decision-making include preoperative decision support, intraoperative assistance functions as well as robot-assisted actions. Examples for such actions along the surgical treatment path are given and discussed in the following.

Preoperative decision support

In medicine, clinical decisions are made on the basis of standardized procedures, evidence-based medicine and individual experience of the physicians. The goal is a personalized therapeutic pathway for each patient, the results of which currently depend largely on the experience of the treating physicians. In surgery, the multitude and variability of diagnostic data and the potentially disruptive nature of major surgical procedures pose special challenges to healthcare professionals as compared to other medical fields. The research field of Surgical Data Science addresses this issue in order to achieve objective patient-specific therapy through data-driven treatment using ML[8].

In this context, decision support systems have two main tasks: balancing the variable level of knowledge and experience of surgeons through objective and patient-specific analysis of patient data and highlighting clinically relevant correlations that are usually not recognized by physicians. Relevant questions can be, for example, if an oncological tumor resection with a required multivisceral resection should be performed at all or if an alternative oncological therapy promises a better outcome in a multimorbid patient with locally advanced cancer.

Currently, work on decision support is mainly limited to risk stratification of surgical treatment or its complications[23]. However, such risk calculations can only be considered as an adjunct to the physician-patient discussion and indication for surgery. On the contrary, ML techniques enable the processing of even more comprehensive patient data sets, which can be transferred into mathematical models and enriched with medical background knowledge to recommend or advise against a specific therapy.

In non-surgical specialties, there are already a large number of medical publications investigating AI in terms of prognosis or therapy recommendations, showing that such systems already have clinical relevance. In particular, there are numerous studies on the superiority of AI-based algorithms over physicians in the evaluation of radiological image data, such as in mammography[24].

In contrast, studies on ML and AI in surgical oncology are mainly limited to the feasibility to predict oncologic outcomes[25], postoperative complications[26], and prediction of mortality and survival[27]. Comparative studies focused on improving treatment decision-making are few[28]. For example, Rice *et al.*[28] investigated survival after esophagectomy with and without neoadjuvant or adjuvant therapy to recommend multimodality therapy. However, data from previous studies were only retrospectively analyzed without prospective validation. Furthermore, successful clinical translation requires the integration of ML algorithms into a user-friendly application that is integrated into the clinical workflow [Figure 1]. As long as the algorithms explored are only in the form of text line programs, they will not be used by physicians and will not help improve treatment.

There have been commercial attempts to address this problem, most famously, IBM Watson for Oncology (WFO), which learns from test cases and experts of the Memorial Sloan Kettering Cancer Center. However,

studies have revealed a low concordance between the WFO treatment recommendations and actual surgical treatments, which may contribute to the lack of clinical trials to date[29]. On the other hand, Google's Deep Mind was criticized for its lack of transparency and remaining problems regarding privacy and power concerns[30]. Both cases highlight the complexity of challenges regarding processing data and the complexity of surgical decision-making in general, which represent a hurdle even for the biggest players in the tech industry. Thus, with respect to "data-driven decision support," there are currently an increasing number of publications linking surgical oncology data with ML. However, the work that actually addresses clinically relevant decision support is scarce, and the integration of AI systems into the clinic is still in its infancy. Such work, however, has the potential to fundamentally change surgical oncology treatment decisions by making them not only evidence-based but also data-based in the future.

Intraoperative assistance
As data monitoring in the modern OR grows more complex, traditional processing techniques may not be able to handle the large amounts of generated data[31]. ML not only allows for the processing of large data sets, but also for combining the different data streams obtained during the treatment of a surgical patient, that is, individual patient data, device data, and case knowledge[7]. The following section gives an overview of the intraoperative use of ML to assist surgeons. It covers the ML guided streamlining of OR management and scheduling to increase efficiency and reduce costs, the recognition of operation phases and surgical workflow as a basis for automated actions and context-aware displays, intraoperative decision support through identification of risk structures or case retrieval, optical biopsy by way of advanced imaging and the potential advantages of XAI solutions.

OR management, including reduction of turnover times, remains a well-researched topic, which is unsurprising as the OR can account for up to 40% of a hospital's costs and 60-70% of its revenue[32]. Current research suggests that more dynamic, parallel work models that overlap workflow steps can increase productivity and reduce turnover times, especially in short-to-medium duration processes, leading to increased cost-effectiveness[33,34]. For example, overlapping patient induction increased 15-40% of operations and OR utilization effectiveness up to 81% in a computer simulation[35], while role definition, task allocation, and sequencing decreased turnover time in robotic surgery[36]. Currently, OR time is assigned by scheduled case duration, which is usually either determined based on previous case duration averages or surgeon estimates. As these methods have limited accuracy, intraoperative monitoring using ML methods to accurately predict case duration could allow for more flexible scheduling and better team organization[37]. In clinical practice, the correct prediction of procedure duration could, for example, ease the coordination of surgical and anaesthesiological decisions like the timing of the last application of muscle relaxant and the reduction of narcotic medication, further shortening turnover times. ML-based models for remaining operation time prediction already exist (e.g., using both video and surgical device data)[38]. Zhao *et al.*[39] compared multiple ML methods, specifically multivariable linear regression, ridge regression, lasso regression, random forest, boosted regression tree, and neural network, resulting in a better performance of all methods than the scheduled case duration.

For most potential intraoperative applications of ML like cognitive camera control[40] or context-aware display of risk structures[41], automated recognition of surgical phases is key to placing the data in its context. Here, surgical phases are defined as higher-level tasks that constitute an entire surgical procedure and can be identified through data inputs like surgical video[42] or instrument use[43]. The ML models used for this purpose are generally based on supervised learning, in which human annotators apply labels to surgical videos, upon which the ML algorithms for computer vision are then trained. In an analysis of 40 laparoscopic sleeve gastrectomies, Hashimoto *et al.*[44] reported 92% accuracy in segmentation and identification of steps when compared to the ground truth of surgical annotation. The authors also pose that

automatic segmentation and labeling of video footage by computer vision may greatly facilitate the training and coaching of young surgeons, as it does not require time-intensive manual viewing and editing.

Because of the importance of annotated data for ML-based workflow recognition, further research into workflow and skill analysis depends on the creation of large, publicly available multicenter data sets like the widely used Cholec80 dataset[45] or the HeiChole Benchmark[46].

While risk stratification usually takes place in the preoperative setting[47], intraoperative surgical decision-making remains a key element contributing to postoperative outcome, frequency of adverse events and patient safety. Human cognitive errors can contribute to up to half of recorded adverse events in surgery and are most commonly observed during the intraoperative phase[48], and decision-making skills may vary with surgical experience[49,50]. In a systematic review not specific to surgery, computerized clinical decision support systems (i.e., diagnostic systems, reminder systems, disease management systems, and drug-dosing or prescribing systems) improved practitioner performance in the majority (64%) of assessed studies[51]. Similarly, the employment of ML in the OR may be able to enhance the surgical decision-making process to allow for fast and accurate responses to patient risks through real-time assistance.

One such approach is the automated identification of risk structures using computer vision. Artificial neural networks have achieved 94.2% accuracy in distinguishing the uterine artery from the ureter in laparoscopic hysterectomy[52], and can be trained to assess the critical view of safety in laparoscopic cholecystectomy[53], which depends on the visual identification of hepatobiliary anatomy before transection to prevent serious adverse events like accidental bile duct injury[54]. Furthermore, approaches for computer vision tasks can be used to improve surgical safety through vigilance in the OR, for example, by tracking the surgical team or analyzing the video stream of the laparoscope, assessment of case difficulty, and behavior analysis of the surgical team[42].

Especially in difficult surgical cases, ML-based systems could provide assistance in the form of case retrieval, whereby surgeons are presented with images or video data from previous cases similar to the operation currently being performed[55].

There are also potential applications for ML to provide an intraoperative pathology assessment by *in vivo* optical biopsy. While *ex vivo* frozen histopathological examination is the current gold standard and is likely to remain as such, several advanced imaging techniques are evolving to supplement the process. This could potentially prove useful in assisting the surgeon's decision whether to biopsy and analyze tissue, provide an intraoperative quality assessment, as well as a preliminary diagnosis when the prompt pathological examination is not possible, as is oftentimes the case on weekends and during night shifts.

Techniques that are currently evolving include probe-based confocal laser endomicroscopy (pCLE), hyperspectral imaging, optical coherence tomography, and contrast-enhanced ultrasound. pCLE is a biophotonics technique that enables direct visualization of tissue at a microscopic level. For example, Li *et al.*[56] used a CNN and a long short-term memory architecture to distinguish between glioblastoma and meningioma. Hyperspectral imaging is the combination of spectroscopy and digital imaging, which analyzes across a wide spectrum of light instead of detecting only red-green-blue colors for each pixel. A modified CNN achieved 92% accuracy in distinguishing thyroid carcinoma from normal tissue on hyperspectral imaging images of *ex vivo* fresh surgical specimens[57]. Optical coherence tomography is a noninvasive cross-sectional imaging technique using low-coherence light. Hou *et al.*[58] used a deep learning approach to identify metastatic lymph nodes and reported 90.1% accuracy. CEUS is an ultrasound-guided

technique in which a hyperechoic intravenously injected contrast agent unmasks areas of heightened perfusion. Ritschel *et al.*[59] applied a support vector machine trained on CEUS image data to detect the resection margins of glioblastoma in neurosurgery.

The majority of the above-mentioned possible applications of ML in the OR are still far from clinical routine use[60]. Part of the hesitation to employ ML in this context may be the rare cases in which ML results in false output due to interpretation of data out of its context[61], which could prove dangerous when applied to patients[62]. Mistrust of ML algorithms may be high, especially in cases of "black box" models, whereby the processes between the input and output of data are not transparent to the health care practitioners using them or the engineers designing them. In contrast, XAI presents the user with the reasons for its predictions, which would allow the surgeons to make independent judgment calls on whether they consider the information reliable. The XAI based warning-system developed by Lundberg *et al.*[63] for hypoxemia during surgical procedures provides a prediction 5 minutes before hypoxemia occurs and lists the incorporated risk factors like comorbidities and vital sign abnormalities in real time. The above-mentioned CEUS guided examination of resection margins in glioblastoma provides color maps for its prediction probabilities[59], allowing surgeons to interpret the results visually.

ML-based systems in the intraoperative setting show promise to markedly improve the OR environment by streamlining workflow processes, supporting decision-making and increasing surgical precision and safety. Much of the current research focuses on technology without immediate clinical applicability[64]. Thus, implementing these concepts still has several challenges to overcome and will require not only XAI, but also a proper interaction design with a focus on clinical workflow integration to realize clinical applications of AI that provide real-time assistance without distracting from the operating field.

Robot-assisted actions
Historically, the first autonomous robotic system, according to Moustris *et al.*[65], was developed by Harris *et al.*[66] and published in 1997. Called Probot, this robot was designed for autonomous transurethral resection of the prostate (TURP) after manual insertion using a resectoscope. Ultrasound scanning was used for the surgeon to plan and visualize the cutting path. In dry lab tests using a plastic model and a potato as a prostate replacement, the system's accuracy was judged similar to human execution and repeatable. Ten TURP procedures were therefore executed on humans successfully. However, the system only automated this procedure-specific task, all the while relying heavily on human planning.

Certainly, autonomous actions in surgery would be cognitive surgical robots that perceive their environment and use ML to perform a context-aware action. In the following, we give an overview of different - preclinical - approaches to automate tasks without tissue contact, such as cameraguidance, or different tissue manipulation tasks such as knot-tying, suturing, grasping, and holding.

To automate camera guidance in laparoscopic surgery, Wagner *et al.*[40] designed a cognitive camera control and used supervised ML to improve the camera guidance quality based on annotations which improved duration as well as the proportion of good guidance quality in a phantom setup. Hutchinson *et al.*[67] trained a CNN to identify the region of interest during robotic surgery, which was evaluated in a dry-lab experiment of pick and place tasks. Rivas-Blanco *et al.*[68] designed a laparoscopic camera similar to a 360° security camera and used ML to track the operator's instruments, tested in an *in vivo* pig experiment.

Regarding autonomous manipulation, a number of works have been published on the automation of knot-tying and suturing. The aim of Mayer *et al.*[69] was for four industrial robot arms to learn to tie surgical

knots. The interface used a recurrent neural network and generated smooth loop trajectories. Padoy and Hager used hidden Markov models to identify the suturing task on the da Vinci telemanipulator (Intuitive Surgical, Sunnyvale, California, USA), and subtasks were automated with a focus on thread-pulling. The transition between manual mode and autonomous control was designed to be seamless and used augmented reality to show the planned paths[70]. Knoll *et al.*[71] used programming by demonstration based on the psychological concept of "scaffolding", that is, the decomposition of complex tasks into simpler task primitives. During task execution, these learned primitives were instantiated with methods from fluid simulation for smooth trajectories in knot-tying on a cadaver heart[71]. Van den Berg *et al.*[72] found that performing multiple iterations of a manually demonstrated action improved the end result of a robotic knot-tying task. The trajectory was optimized using iterative learning to achieve 7-10x speed compared to manual knot-tying[72]. For suturing, Mikada *et al.*[73] trained an autonomous robot grasper with deep learning to identify the master's needle push into tissue. It then estimated when to pull the needle out of the tissue and performed the pull accordingly, correcting its trajectory if failing. It succeeded in a dry lab setting 90% of times after a maximum of three tries[73]. Lu *et al.*[74] used deep learning to segment surgical knot tying data in order to perform the grasping of a suture thread autonomously. It was tested in simulation as well as with the da Vinci Research Kit (Intuitive Foundation, Sunnyvale, California, USA)[74]. Schwaner *et al.*[75] automated the whole actual stitching task successfully with learning from the demonstration on a silicone model. To this end, they constructed an action library of subtasks by learning dynamic movement primitives from human demonstrations and, during execution, registered them to anatomical landmarks using computer vision[75]. Chiu *et al.*[76] made an reinforcement learning (RL) agent to train the regrasping of the needle during suturing in a simulation (97% success rate) as well as in a dry lab experiment (73.3% success rate).

Another manipulation task is grasping and holding the tissue to provide the tissue tension necessary for surgical dissection. With the circle cutting exercise of fundamentals of laparoscopic surgery[77] in mind, Thananjeyan *et al.*[78] used deep RL to train a policy for tensioning a cloth for clean cutting at multiple points of the circle and tested it successfully in a real-world experiment. Nguyen *et al.*[79] (2019) used deep RL to train a policy for tensioning soft tissue to cut forms other than a circle and tested it in a simulation. Pedram *et al.*[80] used RL (Q-Learning) to make a policy to learn tissue manipulation without a model and with limited demonstrator input. The policies were trained and tested in a simulation and got better with more iterations, reaching an average error of less than 10 pixels between the ideal points and the estimated points[80]. Shin *et al.*[81] used RL to train a policy to learn tissue manipulation in a simulation. RL was directly compared to learning from demonstration, and the latter was shown to be the better learning alternative. Also, its technical feasibility was proven in a dry lab experiment with a Raven IV (Applied Dexterity, Seattle, Washington, USA)[81].

Complementing the above-described grasping part of fundamentals of laparoscopic surgery, Xu *et al.*[82] trained a serpentine surgical robot with learning from demonstration to perform the cloth cutting exercise. Here they demonstrated the use of ML to steer serpentine robots, which is otherwise a complicated task due to their multiple degrees of freedom[82].

Apart from that, different tasks have been addressed by single studies. For suctioning of blood, Barragan *et al.*[83] designed an automated suction arm that was proven to reduce the workload of the operator and the task completion time in a phantom bleeding environment with the da Vinci Research Kit. It used deep learning to recognize the bleeding, but the validation was limited to feasibility with $n = 1$. For palpation, Nichols and Okamura manufactured a silicone model with a hard inclusion and used a haptic device to estimate where this inclusion was. Five different modes were tested, with three of them using ML as a basis.

None of them were superior to the manual mode in specificity; however, they outperformed the manual mode in sensitivity[84]. For surgical debridement, Kehoe *et al.*[85] used learning from demonstration and achieved good results on a low-resolution phantom; however, the performance was slower than debridement performed by a human. For needle insertion, Kim *et al.*[86] automated the tool navigation task to insert a needle inside an eye phantom. With ML, the depth of the eye could be estimated by measuring the eyeball's curvature, therefore making the navigation safer than when operated by hand (a 2D goal was enhanced to 3D). Also, the trajectory had less deviation than the mean surgeon tremor, making it superior in accuracy. Similarly, Keller *et al.*[87] trained a policy with learning from demonstration and RL for a corneal needle to be autonomously guided with optical coherence tomography inside human cadaver corneas and showed superior path accuracy compared to manual insertion and path planning only. Baek *et al.*[88] used RL (Q-learning) for collision avoidance path planning in a very simplified cholecystectomy simulation. A policy was trained to move to the gallbladder while avoiding obstacles, achieving a notable learning curve over 5000 iterations in the simulation.

For automating bowel anastomoses, advances to *in vivo* experiments were made recently by Shademan *et al.*[89]. A system called STAR with a KUKA LWR 4+ (KUKA GmbH, Augsburg, Germany) holding an Endo360° suture device (Corporis Medical, Maastricht, Netherlands) used special near-infrared and 3D tracking of specially marked anatomical landmarks for autonomous path planning and execution of sutures. STAR was successfully tested for feasibility in an *in vivo* pig experiment (*n* = 4) and was more consistent in regard to suture spacing than the open, manual laparoscopic and robot-assisted methods in an *ex vivo* porcine experiment. While this work did not make use of ML yet, to execute the same task with the same robotic system complemented by robotic camera guidance, Saeidi *et al.*[90] recently additionally trained a U-Net based CNN. This enabled the system to identify necessary landmarks without path planning markers that also adapted to breathing patterns. The new STAR system yielded encouraging results in regard to consistency and leak-free suturing.

As an assistive approach to human-robot-collaboration, Berthet-Rayne *et al.*[38] presented a framework system on the Raven II, which aimed to combine human contextual awareness with robotic accuracy. They combined a threshold-based task segmentation from demonstrations paired with learning continuous hidden Markov models of these tasks to support movement of the instruments either with haptic guidance or with automated movement. The completion time and traveled distance were superior compared to manual task execution.

Outside of surgery, Ning *et al.*[91] used RL to teach an ultrasound robot how much force to exert on an unknown ultrasound phantom. It then performed a pre-defined trajectory and achieved a "completeness of [...] 98.3%" while using a similar skin area in comparison with a manual approach. However, the stability of the RL agent remained to be improved, especially because of its short training time.

In summary, the translation of autonomous surgical robotics into a clinically relevant setting remains to be seen, illustrated by the fact that most novel approaches are being tested in a simulation or a dry lab environment only, with few making a breakthrough to *in vivo* or cadaver experiments. Meanwhile, the last 2-4 years have also seen a growing number of ML approaches other than learning from demonstration applied to autonomous surgical tasks emerge, which overall had encouraging results. Thus, it is only a matter of time until ML's advancements in sub-tasks add up to fully automated tasks.

DISCUSSION AND CONCLUSION

Data-driven autonomous actions along the surgical treatment path rely on interpreting surgical data with ML methods. However, the integration of AI systems into the clinical workflow to support clinically relevant decisions and improve patient outcomes is still in its infancy and requires a close collaboration of surgeons, computer scientists and industry leaders. Clinical translation from bench to bedside is often associated with a valley of death[92], especially for surgery which poses domain-specific challenges. The challenges regarding the aforementioned application examples are manifold and related to data acquisition, data annotation, data analytics, and aspects related to clinical translation including regulatory constraints apart from ethical aspects. Furthermore, most evaluations have been performed in simulation or dry-lab settings only, especially in robot-assisted actions.

Clinical studies in the field of AI-assisted surgery providing evidence are scarce, as summarized by Maier-Hein *et al.*[7], including a comprehensive list. As stated by Navarrete-Welton *et al.*[64], AI-assisted decision support, in particular for intraoperative assistance, has not unleashed its potential yet and recommends focusing on the generation of representative multi-centric datasets and careful validation regarding the clinical use case to pioneer clinical success stories.

Although the field is progressing slowly, in the future, we can expect that surgical therapy decisions will not only be based on evidence but also on data. We can also expect to see surgical robots not only as passive telemanipulators but also as active assistants.

DECLARATIONS

Authors' contributions
Drafted the initial manuscript text: Wagner M, Bodenstedt S, Speidel S, Daum M, Schulze A, Younis R, and Brandenburg J
Reviewed, edited, and approved the final manuscript: Wagner M, Bodenstedt S, Daum M, Schulze A, Younis R, Brandenburg J, Kolbinger FR, Distler M, Maier-Hein L, Weitz J, Müller-Stich BP, Speidel S

Availability of data and materials
Not applicable.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright
© The Author(s) 2022.

REFERENCES

1.   Nepogodiev D, Martin J, Biccard B, et al. Global burden of postoperative death. *The Lancet* 2019;393:401. DOI PubMed
2.   Weiser TG, Regenbogen SE, Thompson KD, et al. An estimation of the global volume of surgery: a modelling strategy based on available data. *Lancet* 2008;372:139-44. DOI PubMed
3.   Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med* 2002;346:1128-37. DOI PubMed
4.   Tomašev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 2019;572:116-9. DOI PubMed PMC
5.   Muti HS, Heij LR, Keller G, et al. Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. *Lancet Digital Health* 2021;3:e654-64. DOI PubMed PMC
6.   Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115-8. DOI PubMed PMC
7.   Maier-Hein L, Eisenmann M, Sarikaya D, et al. Surgical data science - from concepts toward clinical translation. *Med Image Anal* 2022;76:102306. DOI PubMed
8.   Maier-Hein L, Vedula SS, Speidel S, et al. Surgical data science for next-generation interventions. *Nat Biomed Eng* 2017;1:691-6. DOI PubMed
9.   Russell S, Norvig P. Artificial intelligence, global edition. Available from: https://elibrary.pearson.de/book/99.150005/9781292401171 [Last accessed on 14 Apr 2022].
10. Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey. *Journal of Artificial Intelligence Research* 1996;4:237. DOI
11. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* 2015;61:85-117. DOI PubMed
12. Lecun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. *Neural Computation* 1989;1:541-51. DOI
13. Waibel A, Hanazawa T, Hinton G, Shikano K, Lang K. Phoneme recognition using time-delay neural networks. *IEEE Trans Neural Netw* 1989;37:328-39. DOI
14. Ross T, Zimmerer D, Vemuri A, et al. Exploiting the potential of unlabeled endoscopic video data with self-supervised learning. *Int J Comput Assist Radiol Surg* 2018;13:925-33. DOI PubMed
15. Zhai X, Oliver A, Kolesnikov A, Beyer L. S4L: self-supervised semi-supervised learning. Available from: https://arxiv.org/abs/1905.03670 [Last accessed on 14 Apr 2022].
16. Wang Y, Yao Q, Kwok JT, Ni LM. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput Surv* 2021;53:1-34. DOI
17. Shrestha A, Mahmood A. Review of deep learning algorithms and architectures. *IEEE Access* 2019;7:53040-65. DOI
18. Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision? Available from: https://papers.nips.cc/paper/2017/hash/2650d6089a6d640c5e85b2b88265dc2b-Abstract.html [Last accessed on 14 Apr 2022].
19. Hüllermeier E, Waegeman W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach Learn* 2021;110:457-506. DOI
20. Adler TJ, Ardizzone L, Vemuri A, et al. Uncertainty-aware performance assessment of optical imaging modalities with invertible neural networks. *Int J Comput Assist Radiol Surg* 2019;14:997-1007. DOI PubMed
21. Ardizzone L, Kruse J, Rother C, Köthe U. Analyzing inverse problems with invertible neural networks. Available from: https://arxiv.org/abs/1808.04730 [Last accessed on 14 Apr 2022].
22. Barredo Arrieta A, Díaz-rodríguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 2020;58:82-115. DOI
23. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg* 2018;268:70-6. DOI PubMed PMC
24. Mckinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89-94. DOI PubMed
25. Kudo SE, Ichimasa K, Villard B, et al. Artificial intelligence system to determine risk of T1 colorectal cancer metastasis to lymph

node. *Gastroenterology* 2021;160:1075-1084.e2. DOI PubMed

26. Mai RY, Lu HZ, Bai T, et al. Artificial neural network model for preoperative prediction of severe liver failure after hemihepatectomy in patients with hepatocellular carcinoma. *Surgery* 2020;168:643-52. DOI PubMed

27. Que SJ, Chen QY, Qing-Zhong, et al. Application of preoperative artificial neural network based on blood biomarkers and clinicopathological parameters for predicting long-term survival of patients with gastric cancer. *World J Gastroenterol* 2019;25:6451-64. DOI PubMed PMC

28. Rice TW, Lu M, Ishwaran H, Blackstone EH; Worldwide Esophageal Cancer Collaboration Investigators. Precision surgical therapy for adenocarcinoma of the esophagus and esophagogastric junction. *J Thorac Oncol* 2019;14:2164-75. DOI PubMed PMC

29. Zou FW, Tang YF, Liu CY, Ma JA, Hu CH. Concordance study between IBM watson for oncology and real clinical practice for cervical cancer patients in China: a retrospective analysis. *Front Genet* 2020;11:200. DOI PubMed PMC

30. Powles J, Hodson H. Google deepmind and healthcare in an age of algorithms. *Health Technol (Berl)* 2017;7:351-67. DOI PubMed PMC

31. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013;309:1351-2. DOI PubMed

32. Healey T, El-Othmani MM, Healey J, Peterson TC, Saleh KJ. Improving operating room efficiency, part 1: general managerial and preoperative strategies. *JBJS Rev* 2015;3:e3. DOI PubMed

33. Stahl JE, Sandberg WS, Daily B, et al. Reorganizing patient care and workflow in the operating room: a cost-effectiveness study. *Surgery* 2006;139:717-28. DOI PubMed

34. Marjamaa RA, Torkki PM, Hirvensalo EJ, Kirvelä OA. What is the best workflow for an operating room? *Health Care Manag Sci* 2009;12:142-6. DOI PubMed

35. Bercker S, Waschipky R, Hokema F, Brecht W. [Effects of overlapping induction on the utilization of complex operating structures: estimation using the practical application of a simulation model]. *Anaesthesist* 2013;62:440-6. DOI PubMed

36. Souders CP, Catchpole KR, Wood LN, et al. Reducing operating room turnover time for robotic surgery using a motor racing pit stop model. *World J Surg* 2017;41:1943-9. DOI PubMed PMC

37. Tanzi L, Piazzolla P, Vezzetti E. Intraoperative surgery room management: A deep learning perspective. *Int J Med Robot* 2020;16:1-12. DOI PubMed

38. Berthet-Rayne P, Power M, King H, Yang GZ. Hubot: A three state Human-Robot collaborative framework for bimanual surgical tasks based on learned models. Available from: https://ieeexplore.ieee.org/document/7487198 [Last accessed on 14 Apr 2022].

39. Zhao B, Waterman RS, Urman RD, Gabriel RA. A machine learning approach to predicting case duration for robot-assisted surgery. *J Med Syst* 2019;43:32. DOI PubMed

40. Wagner M, Bihlmaier A, Kenngott HG, et al. A learning robot for cognitive camera control in minimally invasive surgery. *Surg Endosc* 2021;35:5365-74. DOI PubMed PMC

41. Katić D, Wekerle AL, Görtler J, et al. Context-aware Augmented Reality in laparoscopic surgery. *Comput Med Imaging Graph* 2013;37:174-82. DOI PubMed

42. Kennedy-Metz LR, Mascagni P, Torralba A, et al. Computer vision in the operating room: opportunities and caveats. *IEEE Trans Med Robot Bionics* 2021;3:2-10. DOI PubMed PMC

43. Garrow CR, Kowalewski KF, Li L, et al. Machine learning for surgical phase recognition: a systematic review. *Ann Surg* 2021;273:684-93. DOI PubMed

44. Hashimoto DA, Rosman G, Volkov M, Rus DL, Meireles OR. Artificial intelligence for intraoperative video analysis: machine learning's role in surgical education. *Journal of the American College of Surgeons* 2017;225:4. DOI

45. Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Padoy N. EndoNet: A Deep Architecture for Recognition Tasks on Laparoscopic Videos. *IEEE Trans Med Imaging* 2017;36:86-97. DOI PubMed

46. Wagner M, Müller-Stich BP, Kisilenko A, et al. Comparative validation of machine learning algorithms for surgical workflow and skill analysis with the HeiChole benchmark. Available from: https://arxiv.org/ftp/arxiv/papers/2109/2109.14956.pdf.

47. Moonesinghe SR, Mythen MG, Das P, Rowan KM, Grocott MP. Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery: qualitative systematic review. *Anesthesiology* 2013;119:959-81. DOI PubMed

48. Suliburk JW, Buck QM, Pirko CJ, et al. Analysis of human performance deficiencies associated with surgical adverse events. *JAMA Netw Open* 2019;2:e198067. DOI PubMed PMC

49. Flin R, Youngson G, Yule S. How do surgeons make intraoperative decisions? *Qual Saf Health Care* 2007;16:235-9. DOI PubMed PMC

50. Glance LG, Osler TM, Neuman MD. Redesigning surgical decision making for high-risk patients. *N Engl J Med* 2014;370:1379-81. DOI PubMed PMC

51. Garg AX, Adhikari NK, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005;293:1223-38. DOI PubMed

52. Harangi B, Hajdu A, Lampé R, Torok P. Recognizing ureter and uterine artery in endoscopic images using a convolutional neural network. Available from: https://www.semanticscholar.org/paper/Recognizing-Ureter-and-Uterine-Artery-in-Endoscopic-Harangi-Hajdu/b9956febca47a364eb23b6cc65bccfec06206509 [Last accessed on Apr 2022].

53. Mascagni P, Vardazaryan A, Alapatt D, et al. Artificial intelligence for surgical safety: automatic assessment of the critical view of safety in laparoscopic cholecystectomy using deep learning. *Ann Surg* 2020. DOI PubMed

54. Strasberg SM, Brunt LM. Rationale and use of the critical view of safety in laparoscopic cholecystectomy. *J Am Coll Surg*

2010;211:132-8. DOI PubMed

55.   Quellec G, Lamard M, Cazuguel G, et al. Real-time retrieval of similar videos with application to computer-aided retinal surgery. Available from: https://ieeexplore.ieee.org/document/6091107 [Last accessed on 14 Apr 2022].

56.   Li Y, Charalampaki P, Liu Y, Yang GZ, Giannarou S. Context aware decision support in neurosurgical oncology based on an efficient classification of endomicroscopic data. *Int J Comput Assist Radiol Surg* 2018;13:1187-99. DOI PubMed PMC

57.   Halicek M, Little JV, Wang X, et al. Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks. *Proc SPIE Int Soc Opt Eng* 2018;10469:104690X. DOI PubMed PMC

58.   Hou F, Liang Y, Yang Z, et al. Automatic identification of metastatic lymph nodes in OCT images. Available from: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10867/108673G/Automatic-identification-of-metastatic-lymph-nodes-in-OCT-images/10.1117/12.2511588.short?SSO=1 [Last accessed on 14 Apr 2022].

59.   Ritschel K, Pechlivanis I, Winter S. Brain tumor classification on intraoperative contrast-enhanced ultrasound. *Int J Comput Assist Radiol Surg* 2015;10:531-40. DOI PubMed

60.   Maier-Hein L, Eisenmann M, Sarikaya D, et al. Surgical data science - from concepts toward clinical translation. Available from: https://arxiv.org/abs/2011.02284 [Last accessed on 14 Apr 2022].

61.   Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA* 2017;318:517-8. DOI PubMed

62.   Safdar NM, Banja JD, Meltzer CC. Ethical considerations in artificial intelligence. *Eur J Radiol* 2020;122:108768. DOI PubMed

63.   Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* 2018;2:749-60. DOI PubMed PMC

64.   Navarrete-Welton AJ, Hashimoto DA. Current applications of artificial intelligence for intraoperative decision support in surgery. *Front Med* 2020;14:369-81. DOI PubMed

65.   Moustris GP, Hiridis SC, Deliparaschos KM, Konstantinidis KM. Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. *Int J Med Robot* 2011;7:375-92. DOI PubMed

66.   Harris SJ, Arambula-Cosio F, Mei Q, et al. The Probot--an active robot for prostate resection. *Proc Inst Mech Eng H* 1997;211:317-25. DOI PubMed

67.   Hutchinson K, Yasar MS, Bhatia H, Alemzadeh H. A reactive autonomous camera system for the RAVEN II surgical robot. Available from: https://arxiv.org/abs/2010.04785 [Last accessed on 14 Apr 2022].

68.   Rivas-blanco I, Lopez-casado C, Perez-del-pulgar CJ, Garcia-vacas F, Fraile JC, Munoz VF. Smart cable-driven camera robotic assistant. *IEEE Trans Human-Mach Syst* 2018;48:183-96. DOI

69.   Mayer H, Gomez F, Wierstra D, et al. A System for robotic heart surgery that learns to tie knots using recurrent neural networks. Available from: https://ieeexplore.ieee.org/document/4059310 [Last accessed on 14 Apr 2022].

70.   Padoy N, Hager GD. Human-machine collaborative surgery using learned models. Available from: https://ieeexplore.ieee.org/document/5980250 [Last accessed on 14 Apr 2022].

71.   Knoll A, Mayer H, Staub C, Bauernschmitt R. Selective automation and skill transfer in medical robotics: a demonstration on surgical knot-tying. *Int J Med Robot* 2012;8:384-97. DOI PubMed

72.   van den Berg J, Miller S, Duckworth D, et al. Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations. Available from: https://ieeexplore.ieee.org/document/5509621 [Last accessed on 14 Apr 2022].

73.   Mikada T, Kanno T, Kawase T, Miyazaki T, Kawashima K. Suturing support by human cooperative robot control using deep learning. *IEEE Access* 2020;8:167739-46. DOI

74.   Lu B, Chen W, Jin YM, et al. A learning-driven framework with spatial optimization for surgical suture thread reconstruction and autonomous grasping under multiple topologies and environmental noises. Available from: https://arxiv.org/abs/2007.00920 [Last accessed on 14 Apr 2022].

75.   Schwaner KL, Dall'Alba D, Jensen PT, et al. Autonomous needle manipulation for robotic surgical suturing based on skills learned from demonstration. Available from: https://ieeexplore.ieee.org/document/9551569 [Last accessed on 14 Apr 2022].

76.   Chiu ZY, Richter F, Funk EK, et al. Bimanual regrasping for suture needles using reinforcement learning for rapid motion planning. Available from: https://arxiv.org/abs/2011.04813 [Last accessed on 14 Apr 2022].

77.   Peters JH, Fried GM, Swanstrom LL, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery* 2004;135:21-7. DOI PubMed

78.   Thananjeyan B, Garg A, Krishnan S, et al. Multilateral surgical pattern cutting in 2D orthotropic gauze with deep reinforcement learning policies for tensioning. Available from: https://ieeexplore.ieee.org/document/7989275 [Last accessed on 14 Apr 2022].

79.   Nguyen ND, Nguyen T, Nahavandi S, et al. Manipulating soft tissues by deep reinforcement learning for autonomous robotic surgery. Available from: https://ieeexplore.ieee.org/document/8836924 [Last accessed on 14 Apr 2022].

80.   Pedram SA, Ferguson PW, Shin C, et al. Toward synergic learning for autonomous manipulation of deformable tissues via surgical robots: an approximate q-learning approach. Available from: https://arxiv.org/abs/1910.03398 [Last accessed on 14 Apr 2022].

81.   Shin C, Ferguson PW, Pedram SA, et al. Autonomous tissue manipulation via surgical robot using learning based model predictive control. Available from: https://arxiv.org/abs/1902.01459 [Last accessed on 14 Apr 2022].

82.   Xu W, Chen J, Lau HYK, Ren H. Automate surgical tasks for a flexible serpentine manipulator via learning actuation space trajectory from demonstration. Available from: https://ieeexplore.ieee.org/document/7487640 [Last accessed on 14 Apr 2022].

83.   Barragan JA, Chanci D, Yu D, Wachs JP. SACHETS: semi-autonomous cognitive hybrid emergency teleoperated suction. Available from: https://ieeexplore.ieee.org/document/9515517 [Last accessed on 14 Apr 2022].

84. Nichols KA, Okamura AM. A framework for multilateral manipulation in surgical tasks. *IEEE Trans Automat Sci Eng* 2016;13:68-77. DOI

85. Kehoe B, Kahn G, Mahler J, et al. Autonomous multilateral debridement with the Raven surgical robot. Available from: http://people.eecs.berkeley.edu/~pabbeel/papers/2014-ICRA-multilateral-ravens.pdf [Last accessed on 14 Apr 2022].

86. Kim JW, Zhang P, Gehlbach P, et al. Towards autonomous eye surgery by combining deep imitation learning with optimal control. *Proc Mach Learn Res* 2021;155:2347-58. PubMed PMC

87. Keller B, Draelos M, Zhou K, et al. Optical coherence tomography-guided robotic ophthalmic microsurgery via reinforcement learning from demonstration. *IEEE Trans Robot* 2020;36:1207-18. DOI

88. Baek D, Hwang M, Kim H, Kwon D-S. Path planning for automation of surgery robot based on probabilistic roadmap and reinforcement learning. Available from: https://ieeexplore.ieee.org/document/8441801 [Last accessed on 14 Apr 2022].

89. Shademan A, Decker RS, Opfermann JD, Leonard S, Krieger A, Kim PC. Supervised autonomous robotic soft tissue surgery. *Sci Transl Med* 2016;8:337ra64. DOI PubMed

90. Saeidi H, Opfermann JD, Kam M, et al. Autonomous robotic laparoscopic surgery for intestinal anastomosis. *Sci Robot* 2022;7:eabj2908. DOI PubMed PMC

91. Ning G, Chen J, Zhang X, Liao H. Force-guided autonomous robotic ultrasound scanning control method for soft uncertain environment. *Int J Comput Assist Radiol Surg* 2021;16:2189-99. DOI PubMed

92. Butler D. Translational research: crossing the valley of death. *Nature* 2008;453:840-2. DOI PubMed