REVIEW ARTICLE

Machine Learning and Artificial Intelligence in Surgical Fields

Melissa Egert[1] · James E. Steward[1] · Chandru P. Sundaram[1]

Abstract

Artificial intelligence (AI) and machine learning (ML) have the potential to improve multiple facets of medical practice, including diagnosis of disease, surgical training, clinical outcomes, and access to healthcare. There have been various applications of this technology to surgical fields. AI and ML have been used to evaluate a surgeon's technical skill. These technologies can detect instrument motion, recognize patterns in video recordings, and track the physical motion, eye movements, and cognitive function of the surgeon. These modalities also aid in the advancement of robotic surgical training. The da Vinci Standard Surgical System developed a recording and playback system to help trainees receive tactical feedback to acquire more precision when operating. ML has shown promise in recognizing and classifying complex patterns on diagnostic images and within pathologic tissue analysis. This allows for more accurate and efficient diagnosis and treatment. Artificial neural networks are able to analyze sets of symptoms in conjunction with labs, imaging, and exam findings to determine the likelihood of a diagnosis or outcome. Telemedicine is another use of ML and AI that uses technology such as voice recognition to deliver health care remotely. Limitations include the need for large data sets to program computers to create the algorithms. There is also the potential for misclassification of data points that do not follow the typical patterns learned by the machine. As more applications of AI and ML are developed for the surgical field, further studies are needed to determine feasibility, efficacy, and cost.

Keywords Artificial intelligence (AI) · Machine learning (ML) · Artificial neural networks

Introduction to Machine Learning and Artificial Intelligence

The use of artificial intelligence (AI) and machine learning (ML) is rapidly developing within the medical field. Artificial intelligence is the use of software to reproduce or mimic human behavior. Machine learning is an AI technique which utilizes a computer to analyze datasets and learn patterns that can be applied to make conclusions when examining a new data point. One subset of machine learning is deep learning, which uses processing layers to learn multiple levels of abstract patterns and algorithms of data. Deep learning has played a major role in technological advances such as speech recognition and object detection [1]. One of the most popular models of machine learning is artificial neural networks (ANNs). ANNs use inputs and training sets of data to predict outcomes by identifying patterns within the training data [2]. ANNs allow for factors and their respective outcomes to be entered into a computer or machine that will analyze the data and determine which factors are necessary to predict certain outcomes. This allows the system to develop a network of neurons without a human inputting a hypothesis and testing said hypothesis. The patterns will emerge naturally, and the output of the analysis can be used to predict future outcomes.

The objective of this literature review is to explore how artificial intelligence is being used for evaluating and improving surgical skills, diagnosing various surgical pathologies using imaging and tissue specimens, and how surgeons can utilize these technologies in the realm of telemedicine to improve access to care and resources.

✉ Melissa Egert
  mcegert@indiana.edu

✉ Chandru P. Sundaram
  sundaram@iupui.edu

1 Department of Urology, Indiana University School of Medicine, 535 N Barnhill Drive, Suite 150, Indianapolis, IN 46202, USA

Method

A literature review was conducted using search parameters and keywords such as artificial intelligence, machine learning,

deep learning, surgical fields, and surgery. Databases used include PubMed and Cochrane Library.

Evaluation of a Surgeon's Technical Skill

One emerging use of these technologies is assessment of surgical technical skill. Many surgeons are evaluated on outcomes such as complication rates, mortality rates, length of stay, estimated blood loss, patient's length of recovery, and recurrence rates. However, objectively evaluating the technical competence of a surgeon can be difficult. Generally, this is assessed by having research participants review a video of the operation and complete a survey [3, 4]. This can be a time-consuming process requiring the reviewer to watch the video, focus on certain procedural steps, and take note of the technique, movements, and errors that occur. Due to the presence of human error of the observer, this may not be the most reliable method of evaluation. For example, there may be minor motions that go undetected by the human reviewer. Variation between reviewers as well as the subjective nature of the rating experience also limits the review of surgical videos.

Both artificial intelligence and machine learning are being used to assess technical skill in surgery. Computer Vision is a form of machine learning that uses computers to identify objects and patterns in video. A recent study used AI algorithms to identify operative steps in laparoscopic sleeve gastrectomy and found that quantitative data can be obtained from surgical videos with 85.6% accuracy using artificial intelligence [5]. Artificial intelligence is being used to objectively assess surgical skill in a variety of ways. These include the use of electromagnetic sensors attached to instruments, hand-mounted eye trackers, force and torque sensors attached to surgical instruments, and direct capture from the robot [6]. Robotic instrument vibrations can be measured to determine how forcefully the instruments are handled. This new technology can objectively collect data such as the number of times an instrument comes into contact with certain structures. Eye trackers can determine the surgeon's object of focus, which can give information as to the procedural step, what may have been overlooked, and what the subconscious thought process of the surgeon may be. This data can be used to assess a surgeon's skill or experience. For example, experienced surgeons have lower vibration magnitudes and forces, as well as shorter completion times for surgical tasks compared to trainees [7]. Collecting this data and using machine learning algorithms to analyze it provides insight into a surgeon's strengths and weakness. It can help identify which skills and maneuvers are important for good patient outcomes and efficient procedure length.

A 2015 study assessed the cognitive engagement, mental workload, and mental state between novice and expert surgeons during robotic surgery [8]. Surgeons were divided into three groups including beginner, combined competent and proficient, and expert groups based on the Dreyfus model. The surgeons performed basic skills such as ring peg transfer and ball placement, intermediate skills such as suturing and knot tying, and advanced skills such as urethra-vesical anastomosis. The subjects were analyzed using tool-based metrics as well as cognitive-based metrics. Tool-based metrics that were assessed included time to completion, times the camera moved and/or was clutched, as well as errors such as instrument collision and number of times the ball was dropped. Cognitive metrics such as cognitive engagement, mental workload, and mental state were assessed using electroencephalography to monitor brain activity. Significant differences were found between the beginners and experts when performing basic and intermediate skills, as well as number of instrument collisions. Competent, proficient surgeons and expert surgeons differed in terms of cognitive metrics, but not tool-based metrics [8]. Cognitive measures can be utilized as another method to evaluate surgical skills with the help of machine learning to analyze the large datasets.

Surgical technical skill can be evaluated by technologies that are already built into surgical equipment such as the da Vinci Systems recording device. This system can accumulate automated performance metrics (APMs) such as instrument and camera motion. This data can be analyzed using machine learning algorithms to recognize movement patterns that can be used to objectively measure surgical skill and perhaps even predict outcomes [9]. One outcome that is often used to assess surgical skill is the length of the total procedure being performed. Using recordings of laparoscopic or robotic surgeries, machine learning can be implemented to analyze the time is takes to perform critical tasks during the surgery, not simply the overall procedure time. Pauses during the surgery that are considered flow disturbances can be evaluated. Each step of the surgical procedure can be analyzed, and the time it takes to complete various phases of the operation can be compared. Using these algorithms, experienced surgeons can be differentiated from beginners within the first 10 seconds of starting a task with 90% accuracy [10]. Surgical technology such as robotic systems provide valuable data that can be utilized by machine learning algorithms to objectively evaluate a surgeon's technical skill. These algorithms can also detect patterns that lead to better outcomes, which may help in training future surgeons.

Robotic Surgical Training

Surgical training, especially in robotic surgery, is not well-standardized and leads to disparities across training programs [11]. Standardization of robotic surgical training would better assess skill and provide the learner with feedback. One example is the fact that the number and rates of flow disruptions are higher for trainees, which adds to the total length of a case

[12]. Flow disruptions during surgeries, such as issues with team work or external distractions in the OR, are associated with a higher rate of surgical errors [13]. Better preparation for surgical cases can limit the flow disruptions. Machine learning algorithms can be used to assess a trainee's success in completing these tasks prior to the OR and provide them with objective feedback to improve. The da Vinci training simulations can further the skills and preparedness of trainees. In the future, robotic surgery curricula may include the option for trainees to record their sessions and collect data on movements. ML algorithms can assess the parts of the procedure that need to be improved on in order to successfully and efficiently perform the procedure. A curriculum should include a way for the learner to demonstrate competence in performing robotic skills and surgical tasks prior to independent practice [14]. Trainees can reach benchmarks and move onto more difficult tasks when the basic skill is completed with minimal error.

Many of the current robotic surgery technologies lack the ability to provide haptic feedback during a case [15]. Haptic feedback, such as the amount of force being applied, and sensory feedback are both important to learning. The da Vinci Standard Surgical System developed a recording and playback system, which is a platform that can be utilized to better train surgeons [16]. A learner can watch recorded surgical procedures and feel recorded movements of the controls and variable speeds. Using this branch of artificial intelligence, novices can have tactile feedback of the correct motions to perform during procedures or various surgical tasks. Therefore, if a learner is struggling with a specific task or motion, feeling the correct movements may help the trainee learn the technique more quickly. Although further developments in technology are needed in order to provide real-time feedback during operations, the recording and playback system is a step in the right direction in helping develop muscle memory for trainees.

Use in Diagnosis and Workup

Artificial intelligence is beginning to play a major role in the use of imaging for diagnosis in various medical fields. Machine learning can recognize complex patterns on various radiologic modalities such as CT, MRI, and PET images to support a radiologist's judgment and clinical decision making [17]. Machine learning can be used to create algorithms to detect normal versus pathologic tissue on imaging by using medical image segmentation. It also aids in orienting different imaging modalities such as an ultrasound, MRI, and CT of the kidney into a common framework so the information from these different studies can be paralleled or combined. Imaging is an important diagnostic step in multiple surgical specialties. The use of AI in combination with radiologists could aid surgeons in developing a more accurate surgical plan before the patient is on the operating table.

A rapidly developing use for artificial intelligence is in the detection of cancer. Artificial intelligence can help radiologists make interpretations about the likelihood that a certain lesion is cancerous based on past patterns and diagnoses. AI can be used to help delineate the volume of cancers, monitor growth over time, and predict the biologic course and clinical outcome based on the radiologic phenotype [18]. This information can guide next steps and give the patient and the physician more information to develop a treatment plan. A surgeon could use this information when deciding which malignancies are operable. For example, if algorithms can determine likelihood that malignancy is present when analyzing an MRI of the prostate with a patient with borderline PSA, a more informed decision can be made to proceed with biopsy or continue active surveillance.

Artificial neural networks have been used to create algorithms for diagnosis and management of medical conditions. Various structures of ANNs have been shown to be superior in diagnosing appendicitis compared to clinical scoring systems that have previously been used [19]. For example, input variables such as pain location, rebound tenderness, bowel sounds, nausea, WBC count, and tenderness of the RLQ can be used, and the likelihood of acute appendicitis then generated [19]. Surgeons can use ANNs to help guide their diagnosis. ANNs have been proven to increase radiologists' accuracy in differentiating between malignant and benign pulmonary nodules on CT [20]. In this study, ANNs were used to distinguished if a nodule was benign or malignant based on seven clinical parameters and sixteen radiologic findings, and they were presented with the data both with and without the ANN assistance [20]. ANNs can give practitioners additional information to guide their decision making and recommendations.

Pathology is another field that has benefitted from the advances in machine learning and artificial intelligence. When biopsies are taken by a surgeon, there is often a small focus of malignancy within a large amount of benign tissue. Looking for a malignancy in this situation is a time-consuming process for pathologists, and a small focus of malignancy may be overlooked. Deep learning could improve the objectivity of making the diagnosis by allowing pathologists to focus on certain portions of the tissue that have a higher probably of malignancy. If more efficient and accurate recommendations can be made by pathologists, surgeons can more rapidly make a diagnosis and treatment plan. In fact, deep learning has been shown to improve the accuracy of diagnosing prostate cancer and detecting lymph node metastasis in patients with breast cancer [21]. This allows surgeons to make a more accurate plan when deciding whether to allow a patient to undergo active surveillance versus an operation or whether to extend a lymph node dissection during an operation. Multiple histologic slides that had been previously read by pathologists were

used as the input for a deep learning network. ANNs are then developed from the input and can be used to analyze subsequent biopsies.

Telemedicine

Another direction of machine learning and artificial intelligence within the medical field is telemedicine. Telemedicine uses technology to deliver care from afar. Physicians can remotely track patients' progress or connect patients with resources. This may help in rural areas with less access to healthcare and could potentially lower the overall costs of healthcare. Telemedicine has been studied for use in patient education, monitoring of chronic conditions, screening for disease, assessment of clinical presentations, and treatment [22]. Telemedicine can also be used for understaffed, underfunded hospitals, which lack access to certain healthcare providers. Theoretically, one physician at a medical hub could monitor several patients in different locations and work with the team present with a particular patient to provide care [23].

Voice recognition technology is being used to help patients input their daily weights, symptoms, blood pressures, diet, exercise, and medication adherence into the telemedicine system so the healthcare team can follow patients more closely. Information Communication technology based telehealth programs with voice recognition are shown to improve sodium intake and quality of life in chronic heart failure patients who utilized the program appropriately [24]. Additionally, telemedicine has been found to be as effective as face-to-face interactions or telephone calls for managing heart failure [22]. Despite this, the cost of telemedicine is still controversial and depends on the type of care being provided [23]. The effectiveness of telemedicine in reducing mortality is still debated.

Limitations and Future Directions

Currently, machine learning and artificial intelligence can assist in training basic surgical skills. Further algorithms with increasing complexity need to be developed to train surgeons to complete more complicated procedures. Educators will still be necessary to teach trainees certain techniques and the indications for these procedures. Additionally when evaluating skills, the parameters by which skill is evaluated should be agreed upon. For instance, the amount of time it takes to complete a surgical task may not always correlate to the task being performed correctly. This is one piece of data to consider in addition to factors such as the degree of difficulty of the procedure, intraoperative complications, and clinical outcome.

One major limitation of machine learning is that creating the algorithms relies on a set of input data. The larger the input, the better the pattern recognition and application of

the learning. For less common disease processes that do not follow a distinct pattern, machine learning may be difficult to utilize. Additionally, if the case that is being analyzed by the algorithm is an exception to the usual presentation, a diagnosis could be missed. This highlights the importance of using data obtained from machine learning only as another piece of information to aid in making a diagnosis. Clinical judgment is still necessary to make a diagnosis and create a treatment plan.

In regard to telemedicine, further analysis of cost and mortality benefit need to be studied to determine the value of its use. Individual hospitals need to weigh the costs and benefits of using systems such as ICU telemedicine. Physicians and hospital systems would also have to decide how beneficial the use of telemedicine would be for their patient population and scope of practice. Using telemedicine to track patient's progress or compliance may come with a learning curve for the health care staff as well as the patient. Patients and physicians would both need to be willing to learn and use the system.

In summary, machine learning and artificial intelligence are quickly being incorporated into multiple aspects of medicine related to surgical fields. This technology can benefit surgeons when evaluating surgical skills as well as teaching trainees the best methods while operating, especially during robotic procedures. Machine learning can also be implemented to improve diagnosis of certain conditions and cancers both radiographically and from a pathology standpoint. Combining artificial intelligence with the clinical judgment of a physician can lead to better outcomes and more efficient shared decision-making. Telemedicine is another aspect of artificial intelligence that is aiding in delivering care to underserved health care communities. Further study of the benefits and limitations of this technology is vital to its incorporation into modern healthcare in a safe and effective manner.

Compliance with Ethical Standards

References

1. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature. 521(7553):436–444. https://doi.org/10.1038/nature14539

2. Lundervold AS, Lundervold A (2018) An overview of deep learning in medical imaging focusing on MRI. Zeitschrift fur Medizinische Physik 29:102–127. https://doi.org/10.1016/j.zemedi.2018.11.002

3. Vaughn CJ, Kim E, O'Sullivan P, Huang E, Lin MYC, Wyles S et al (2016) Peer video review and feedback improve performance in basic surgical skills. Am J Surg 211(2):355–360. https://doi.org/10.1016/j.amjsurg.2015.08.034

4. Nakada SY, Hedican SP, Bishoff JT, Shichman SJ, Wolf JS Jr (2004) Expert videotape analysis and critiquing benefit laparoscopic skills training of urologists. JSLS. 8(2):183–186

5. Hashimoto DA, Rosman G, Witkowski ER, Stafford C, Navarette-Welton AJ, Rattner DW, Lillemoe KD, Rus DL, Meireles OR (2019) Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. Annals of Surg 270:414–421. https://doi.org/10.1097/SLA.0000000000003460

6. Vedula SS, Ishii M, Hager GD (2017) Objective assessment of surgical technical skill and competency in the operating room. Annu Rev Biomed Eng 19(1):301–325. https://doi.org/10.1146/annurev-bioeng-071516-044435

7. Gomez ED, Aggarwal R, McMahan W, Bark K, Kuchenbecker KJ (2016) Objective assessment of robotic surgical skill using instrument contact vibrations. Surg Endosc 30:1419–1431. https://doi.org/10.1007/s00464-015-4346-z

8. Guru KA, Esfahani ET, Raza SJ, Bhat R, Wang K, Hammond Y, Wilding G, Peabody JO, Chowriappa AJ (2015) Cognitive skills assessment during robotic-assisted surgery: separating the wheat from the chaff. BJU Int 115(1):166–174. https://doi.org/10.1111/bju.12657

9. Hung AJ, Chen J, Gill IS (2018) Automated performance metrics and machine learning algorithms to measure surgeon performance and anticipate clinical outcomes in robotic surgery. JAMA Surg 153(8):770–771. https://doi.org/10.1001/jamasurg.2018.1512

10. French A, Lendvay TS, Sweet RM, Kowalewski TM (2017) Predicting surgical skill from the first N seconds of a task: value over task time using the isogony principle. Int Jour of Comp Assis Rad and Surg 12(7):1161–1170. https://doi.org/10.1007/s11548-017-1606-5

11. Carpenter BT, Sundaram CP (2017) Training the next generation of surgeons in robotic surgery. Robot Surg 4:39–44. https://doi.org/10.2147/RSRR/S70552

12. Jain M, Fry BT, Hess LW, Anger JT, Gewertz BL, Catchpole K (2016) Barriers to efficiency in robotic surgery: the resident effect. J Surg Res 205(2):296–304. https://doi.org/10.1016/j.jss.2016.06.092

13. Wiegmann DA, ElBardissi AW, Dearani JA, Daly RC, Sundt TM (2007) Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. Surgery. 142(5):658–665. https://doi.org/10.1016/j.surg.2007.07.034

14. Sridhar AN, Briggs TP, Kelly JD, Nathan S (2017) Training in robotic surgery—an overview. Current Urology Reports.:18–18. https://doi.org/10.1007/s11934-017-0710-y

15. Okamura AM (2009) Haptic feedback in robot-assisted minimally invasive surgery. Curr Opin Urol 19(1):102–107. https://doi.org/10.1097/MOU.0b013e32831a478c

16. Pandya A, Eslamian S, Ying H, Nokleby M, Reisner LA (2019) A robotic recording and playback platform for training surgeons and learning autonomous behaviors using the da Vinci Surgical system. Robotics 8(9). https://doi.org/10.3390/robotics8010009

17. Wang S, Summers RM (2012) Machine learning and radiology. Med Image Anal 16(5):933–951. https://doi.org/10.1016/j.media.2012.02.005

18. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A et al (2019) Artificial intelligence in cancer imaging: clinical challenges and applications. CA A Cancer J Clin 69:127–157. https://doi.org/10.3322/caac.21552

19. Park SY, Kim SM (2015) Acute appendicitis diagnosis using artificial neural networks. Technol Health Care 23(Suppl 2):S559–S565. https://doi.org/10.3233/THC-150994

20. Matsuki Y, Nakamura K, Watanabe H, Aoki T, Nakata H, Katsuragawa S, Doi K (2002) Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution CT evaluation with receiver operating characteristic analysis. Am J Roentgenol 178:657–663. https://doi.org/10.2214/ajr.178.3.1780657

21. Litjens G, Sanchez CI, Timofeeva N, Hermsen M, Nagtegaal I, Kovacs I et al (2016) Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci Rep 6:26286. https://doi.org/10.1038/srep26286

22. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S (2015) Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015(9):CD002098. https://doi.org/10.1002/14651858.CD002098.pub2

23. Kahn JM, Rak KJ, Kuza CC, Ashcraft LE, Barnato AE, Fleck JC et al (2019) Determinants of intensive care unit telemedicine effectiveness. An ethnographic study. Am J Respir Crit Care Med 199(8):970–979. https://doi.org/10.1164/rccm.201802-0259OC

24. Lee H, Park JB, Choi SW, Yoon YE, Park HE, Lee SE, Lee SP, Kim HK, Cho HJ, Choi SY, Lee HY, Choi J, Lee YJ, Kim YJ, Cho GY, Choi J, Sohn DW (2017) Impact of a telehealth program with voice recognition technology in patients with chronic heart failure: feasibility study. JMIR Mhealth Uhealth 5(10):e127. https://doi.org/10.2196/mhealth.7058