Artificial Intelligence Surgery

Most surgeons are skeptical as to the feasibility of autonomous actions in surgery. Interestingly, many examples of autonomous actions already exist and have been around for years. Since the beginning of this millennium, the field of artificial intelligence (AI) has grown exponentially with the development of machine learning (ML), deep learning (DL), computer vision (CV) and natural language processing (NLP). This entry will highlight the most recent issues regarding how AI will get us to more autonomous actions in surgery by discussing the different degrees of surgical autonomy, recent advances with reinforcement learning and the ethical roadblocks that lie ahead.

1. Introduction

Unlike artificial intelligence (AI) in medicine, which hopes to use autonomously functioning algorithms to better analyze patient data in an effort to improve patient outcomes, AI in surgery also involves movement. Unlike strict medicine, surgery is an art that is also dynamic. When we use the term "surgery" we are also referring to endoscopy and interventional techniques and procedures, because interventional disciplines continue to coalesce into the same field, a trend that is seen by the continued increase in hybrid operating rooms that have angiography compatible tables, mobile CT scanners, minimally invasive surgical equipment and endoscopes all in the same room that can be used in tandem. Because of the fact that interventional fields of medicine also rely greatly on the medical management of patients, we believe that AI medicine (AIM) and AI surgery (AIS) could one day be considered two distinct disciplines, with AIM reserved for instances where computer algorithms are used to better diagnose, manage or treat patients without a specific interventional procedure being done.

AIS could be a term for autonomously acting machines that can do interventional gestures/actions. According to the Gartner Hype Cycle, many surgeons believe that we are languishing in the "Trough of Disillusionment", and the promise of autonomous surgery seems like a "pipe dream" for most modern-day surgeons [1][2]. However, the reality is that instances of autonomous actions in surgery already exist. Unfortunately, the reluctance of many laparoscopic surgeons to give up on haptics, or the sense of touch, is actually hindering progress in AIS because of the refusal to embrace robotic tele-manipulation technology, in effect, they are refusing to let go, something that will be needed if the dream of AIS is ever to come to pass [3]. Unfortunately, the medical community has already been shown to be resistant to any automation of medical tasks even simple computations. It is safe to say that automation of surgical tasks will have an even more profound degree of resistance [4].

Another obstacle to the growth of AIS is the dogmatic belief that for something to have AI, it must have algorithms that enable progressive improvement and learning by an artificially intelligent device [5][6]. This creates a conundrum, as theoretically, machine learning (ML) should be infinite, and because of this one wonders what the ultimate purpose of perpetual learning is in surgery. Should it be carried out in the hopes that ultimately the surgical action will become so perfect that it is no longer necessary? How much more perfectly does a robot need to place a titanium clip on the cystic duct?

It could be argued that technology used to create monopolar and bipolar technology is an example of AIS as it has tissue sensing technology that adjusts the action of cautery based on the resistance and capacitance of the tissue (TissueFect™, ValleyLab, Medtronic, Dublin, Ireland). In particular, it has a closed-loop control that analyzes 434,000 data points per second. Does this technology need to improve on that level of data analysis to be considered AI? Or what about Ligasure technology, which uses a high-speed advanced algorithm to seal vessels with changes to the duration of action dependent on tissue thickness (Ligasure, ValleyLab, Medtronic, Dublin, Ireland). We certainly do not mean to imply that there is no room for improvement in these technologies, but at what point should something be defined as AI? Shouldn't any autonomous action be acknowledged and celebrated as an example of AIS?

2. Machine Learning

Machine learning (ML) is a genre of artificial intelligence including algorithms that allow machines to solve problems without specific computer programing. While analyzing big data, machines are enabled to assimilate a large amount of information, applicable for risk stratification, diagnosis, treatment decisions, and survival predictions. Not only can AI models analyze large amounts of data collected over long periods of time, providing predictions for future events on the basis of the statistical weight of past correlations, they can also continuously improve with new data. Through a process called "incremental learning", trainable neural networks improve over time, surpassing unchanging scoring systems and standardized software. Moreover, the human–machine interaction further improves the performance of ML tools. Indeed, the learning process goes far beyond the textbook data, incorporating real-life scenarios and can improve experts' opinions.

Most of the studies conducted on ML tools have focused on machine vision, biomarker discovery, and clinical matching for diagnosis, classification and outcome prediction [7]. Several studies have applied different ML tools to surgery and, in particular, to risk assessment, performance evaluation, treatment decision making and outcome prediction. In an effort to better identify high-risk surgical patients from complex data, a ML project trained on Pythia was built by Corey et al. to predict postoperative complication risk [8]. By using surgical patient electronic health record (EHR) data, including patient demographics, smoking status, medications, co-morbidities, surgical details, and variables addressing surgical complexity, the authors created a decision support tool for the identification of high-risk patients. Similarly, Bertsimas et al. applied novel ML techniques to design an interactive calculator for emergency surgery [9]. By using data of the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database, the authors designed the POTTER application (Predictive OpTimal Trees in Emergency Surgery Risk), to predict postoperative mortality and morbidity [9]. Differently from the standard predictive models, POTTER accounted for nonlinear interactions among variables; thus, the system reboots after each answer, interactively dictating the subsequent question. A similar tool, predicting eight specific postoperative complications and death within one year after major surgery, was developed and validated by Bihorac et al. [10]. MySurgeryRisk can accurately predict acute kidney injury, sepsis, venous thromboembolism, intensive care unit admission >48 h, mechanical ventilation >48 h, and wound, neurologic, and cardiovascular complications with AUC values ranging between 0.82 and 0.94 and death up to 24 months after surgery with AUC values ranging between 0.77 and 0.83. The following studies built different ML models predicting postoperative outcomes that proved to perform better than traditional risk calculators [11][12].

In an attempt to better define patient outcomes, Hung et al. applied ML to objectively measure surgeon performance in a pilot study [13]. By combining procedure-specific automated performance metrics with ML algorithms the authors were able to objectively assess surgeon's performance after robotic surgery. These algorithms still include biases and misinformation, and thus multi-institutional data from several high-volume groups are needed to train the tool and to create a robust model that is able to correctly predict different clinical outcomes.

Surgical performance has also been measured to allow surgeons to learn from their experience and refine their skills. Either by using surgical video clips or by applying virtual reality scenarios, deep learning models were trained to estimate performance level and specific surgical skills [14][15]. In the near future, we will be able to personalize training through ML tools such as these.

Other studies have used ML for prediction of staging [16], for treatment decision making [17], and to improve case duration estimations [18], further expanding the applicability of ML. Despite these advantages, ML presents several challenges, such as the need to process a large amount of data before it can be analyzed, the necessity of repetitively training the model, and of refining it according to the various clinical scenarios. Ethical considerations should also be taken into account when applying ML to healthcare, including privacy and security issues, and the risk for medico-legal implications. These issues will be discussed at more length below [7].

3. Natural Language Processing

Natural Language Processing (NLP) is the evolution of the interaction of Artificial Intelligence (AI) and linguistics. Distinct from simple text information retrieval (IR), which indexes and searches a large volume of text based on a statistical technique [19], NLP evolved from basic approaches (word to word), through a complex process of coding words, sentences, meanings and contexts, to its modern structure. Since 1950, NLP and IR converged into what is known today as NLP, namely, a computer-based algorithm that handles and elaborates natural (human) language, making it suitable for computation [20].

When applied to healthcare, where available clinical data are kept in Electronic Health Records (EHRs), the need to decode a narrative text coming from this large amount of unstructured data has become urgent because of the complexity of the human language and the routine employment of metaphors and telegraphic prose. When compared to NLP, manual reviewing of EHRs is time consuming, possibly misleading because of biases, and extremely expensive [21].

The tremendous potential value of a big data analytical system in healthcare can be easily explained: EHRs represent at this time the major source of patient information, but unfortunately, for the most part, data regarding symptoms, risk factors for a specific disease or outcomes after medical or surgical treatment come from unstructured text. The ability to translate this amount of information into a coded algorithm could allow for more precise screening programs and modify medical and/or surgical strategies. A systematic review from Kolech et al. analyzed the available methods, employing NLP to interpret symptoms from EHRs of inpatients and outpatients, finding possible future applications for NLP in the normalization of symptoms to controlled vocabularies, in order to avoid overlapping of different words for the same concept [21]. A notable criticism of the available studies has been that reported signs and symptoms are easily mixed as the same variable, making interpretation confusing. In this review, only 11% of studies focused on cancer patients, in contrast with the fact that, currently, a major area of interest for AI (not only NLP) is oncology, where early detection of cancer-specific symptoms could facilitate early diagnosis and potentially enhance screening techniques.

An obvious and immediate advantage of having reliable and decoded data coming from clinical notes is the positive impact on the quality of retrospective studies. Moreover, NLP analysis of symptoms and signs in cancer patients may allow for the improved definition of prognostic factors other than surgical and oncological parameters [22]. Emotional status and quality of life of patients after cancer surgery or other cancer treatment has also been investigated through NLP [23] [24]. Banerjee et al., with the creation of a specific domain vocabulary of side-effects after prostatectomy, were able to evaluate minor outcomes hidden in clinical free text, resulting in better management, which could be a game-changer in a population with a 5-year life expectancy rate approaching 99% [23].

When applied to surgery, NPL has been extensively proposed pre-operatively and looking at different post-operative complications such as surgical site infection (SSI). Bucher et al. developed and validated a model of SSI prediction with NLP algorithm by analyzing EHRs from 21,000 patients entered into the ACS-NSQIP, using only text-based documents regarding history and physical condition, operative, progress and nursing notes, radiology reports and discharge summaries [25]. This predictive model had a sensitivity of 79% and a specificity of 92% on external validation, but its added value was the absolute reliable negative predictive value (NPV), which is a relevant issue for events with a low incidence. Anastomotic leak [26], deep venous thrombosis, myocardial infarction and pulmonary embolism were also frequently investigated and results from a recent meta-analysis [27] demonstrated that performance of NLP methods in detection of postoperative complications is similar, if not superior, to non-NLP methods, with a sensitivity of 0.92 versus 0.58 ($p < 0.001$), and comparable specificity. Moreover, NLP models seem to be better than non-NPL models for ruling out specific surgical outcomes, owing to an optimal true-negative identification power. Interestingly, the ability of algorithms to self-correct can increase the utility of their predictions as datasets grow to become more representative of a patient population.

These NLP applications are surely beneficial for patient management, providing better understanding of peri-operative data, but it can be a useful tool for surgeons as well, particularly when applied to surgical education. For example, decoding intra-operative dialogues between residents and faculty, combining NLP and CV, can create and implement a dataset of technique and surgical strategy, thus, creating a real-life textbook of surgery. In addition, NLP has been efficiently used [28] to assess Entrustable Professional Activities (EPAs). EPAs describe specific behaviors associated with different training levels of residents and it can potentially enhance the understanding of their training and autonomy in surgical practice.

NLP can be used to validate datasets that are the basis of surgical risk predictive models, but the main limit of their widespread use is the non-homogeneity of NLP models and EHRs data entry forms across institutions and countries. A future improvement would entail expansion of registries from local to national and international levels to set algorithms that can be externally validated on various populations. Surgeons and their low-level confidence with AI represent another limit to developing a system that is theoretically perfect and promising: it is important for them to understand how AI may impact healthcare and to elaborate strategies of safe interaction to implement this nascent technology. Synergy between fields of AI is also essential in expanding its applications. Lastly, ethical issues and privacy rules protecting patients' sensitive data, can limit the large-scale applicability of NLP over EHRs. Nonetheless, the enormous potential of NLP remains fascinating and the multiple potential benefits of its integration into healthcare must be balanced with risks. Although the technology does not currently exist for NLP to influence autonomous actions in surgery, it must be remembered that communication among team members during surgery is fundamental to the successful performance of

surgery. Additionally, devices already exist and are used today that are voice-activated (ViKY, Endocontrol, Grenoble, France), and it is conceivable that NLP could eventually evolve to benefit the action of voice-controlled devices during a procedure [29][30].

References

1. Oosterhoff, J.H.; Doornberg, J.N. Machine Learning Consortium Artificial intelligence in orthopaedics: False hope or not? A narrative review along the line of Gartner's hype cycle. EFORT Open Rev. 2020, 5, 593–603.

2. Gumbs, A.A.; Perretta, S.; D'Allemagne, B.; Chouillard, E. What is Artificial Intelligence Surgery? Artif. Intell. Surg. 2021, 1, 1–10.

3. Gumbs, A.A.; De Simone, B.; Chouillard, E. Searching for a better definition of robotic surgery: Is it really different from laparoscopy? Mini Invasive Surg. 2020, 2020, 90.

4. Randolph, A.G.; Haynes, R.B.; Wyatt, J.C.; Cook, D.J.; Guyatt, G.H. Users' Guides to the Medical Literature: XVIII. How to use an article evaluating the clinical impact of a computer-based clinical decision support system. JAMA 1999, 282, 67–74.

5. Kassahun, Y.; Yu, B.; Tibebu, A.T.; Stoyanov, D.; Giannarou, S.; Metzen, J.H.; Poorten, E.V. Surgical robotics beyond enhanced dexterity instrumentation: A survey of machine learning techniques and their role in intelligent and autonomous surgical actions. Int. J. Comput. Assist. Radiol. Surg. 2015, 11, 553–568.

6. Hashimoto, D.A.; Rosman, G.; Rus, D.; Meireles, O.R. Artificial Intelligence in Surgery: Promises and Perils. Ann. Surg. 2018, 268, 70–76.

7. Ngiam, K.Y.; Khor, I.W. Big data and machine learning algorithms for health-care delivery. Lancet Oncol. 2019, 20, e262–e273.

8. Corey, K.M.; Kashyap, S.; Lorenzi, E.; Lagoo-Deenadayalan, S.A.; Heller, K.; Whalen, K.; Balu, S.; Heflin, M.T.; McDonald, S.R.; Swaminathan, M.; et al. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data (Pythia): A retrospective, single-site study. PLoS Med. 2018, 15, e1002701.

9. Bertsimas, D.; Dunn, J.; Velmahos, G.C.; Kaafarani, H.M.A. Surgical Risk Is Not Linear: Derivation and Validation of a Novel, User-friendly, and Machine-learning-based Predictive OpTimal Trees in Emergency Surgery Risk (POTTER) Calculator. Ann. Surg. 2018, 268, 574–583.

10. Bihorac, A.; Ozrazgat-Baslanti, T.; Ebadi, A.; Motaei, A.; Madkour, M.; Pardalos, P.M.; Lipori, G.; Hogan, W.R.; Efron, P.A.; Moore, F.; et al. MySurgeryRisk: Development and Validation of a Machine-learning Risk Algorithm for Major Complications and Death After Surgery. Ann. Surg. 2019, 269, 652–662.

11. Chiew, C.J.; Liu, N.; Wong, T.H.; Sim, Y.E.; Abdullah, H.R. Utilizing Machine Learning Methods for Preoperative Prediction of Postsurgical Mortality and Intensive Care Unit Admission. Ann. Surg. 2020, 272, 1133–1139.

12. El Hechi, M.W.; Eddine, S.A.N.; Maurer, L.; Kaafarani, H.M. Leveraging interpretable machine learning algorithms to predict postoperative patient outcomes on mobile devices. Surgery 2021, 169, 750–754.

13. Hung, A.J.; Chen, J.; Gill, I.S. Automated Performance Metrics and Machine Learning Algorithms to Measure Surgeon Performance and Anticipate Clinical Outcomes in Robotic Surgery. JAMA Surg. 2018, 153, 770–771.

14. Winkler-Schwartz, A.; Yilmaz, R.; Mirchi, N.; Bissonnette, V.; Ledwos, N.; Siyar, S.; Azarnoush, H.; Karlik, B.; Del Maestro, R. Machine Learning Identification of Surgical and Operative Factors Associated With Surgical Expertise in Virtual Reality Simulation. JAMA Netw. Open 2019, 2, e198363.

15. Khalid, S.; Goldenberg, M.; Grantcharov, T.; Taati, B.; Rudzicz, F. Evaluation of Deep Learning Models for Identifying Surgical Actions and Measuring Performance. JAMA Netw. Open 2020, 3, e201664.

16. O'Sullivan, B.; Huang, S.H.; De Almeida, J.R.; Hope, A. Alpha Test of Intelligent Machine Learning in Staging Head and Neck Cancer. J. Clin. Oncol. 2020, 38, 1255–1257.

17. Maubert, A.; Birtwisle, L.; Bernard, J.; Benizri, E.; Bereder, J. Can machine learning predict resecability of a peritoneal carcinomatosis? Surg. Oncol. 2019, 29, 120–125.

18. Bartek, M.A.; Saxena, R.C.; Solomon, S.; Fong, C.T.; Behara, L.D.; Venigandla, R.; Velagapudi, K.; Lang, J.D.; Nair, B.G. Improving Operating Room Efficiency: Machine Learning Approach to Predict Case-Time Duration. J. Am. Coll. Surg. 2019, 229, 346–354.e3.

19. Nadkarni, P.M.; Ohno-Machado, L.; Chapman, W. Natural language processing: An introduction. J. Am. Med. Inform. Assoc. 2011, 18, 544–551.

20. Yim, W.W.; Yetisgen, M.; Harris, W.P.; Kwan, S.W. Natural Language Processing in Oncology: A Review. JAMA Oncol. 2016, 2, 797–804.

21. Koleck, T.A.; Dreisbach, C.; Bourne, P.E.; Bakken, S. Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review. J. Am. Med. Inform. Assoc. 2019, 26, 364–379.

22. Hughes, K.S.; Zhou, J.; Bao, Y.; Singh, P.; Wang, J.; Yin, K. Natural language processing to facilitate breast cancer research and management. Breast J. 2019, 26, 92–99.

23. Banerjee, I.; Li, K.; Seneviratne, M.; Ferrari, M.; Seto, T.; Brooks, J.D.; Rubin, D.L.; Hernandez-Boussard, T. Weakly supervised natural language processing for assessing patient-centered outcome following prostate cancer treatment. JAMIA Open 2019, 2, 150–159.

24. Zunic, A.; Corcoran, P.; Spasic, I. Sentiment Analysis in Health and Well-Being: Systematic Review. JMIR Med. Inform. 2020, 8, e16023.

25. Bucher, B.T.; Shi, J.; Ferraro, J.P.; Skarda, D.E.; Samore, M.H.; Hurdle, J.F.; Gundlapalli, A.V.; Chapman, W.W.; Finlayson, S.R.G. Portable Automated Surveillance of Surgical Site Infections Using Natural Language Processing: Development and Validation. Ann. Surg. 2020, 272, 629–636.

26. Soguero-Ruiz, C.; Hindberg, K.; Rojo-Alvarez, J.L.; Skrovseth, S.O.; Godtliebsen, F.; Mortensen, K.; Revhaug, A.; Lindsetmo, R.-O.; Augestad, K.M.; Jenssen, R. Support Vector Feature Selection for Early Detection of Anastomosis Leakage from Bag-of-Words in Electronic Health Records. IEEE J. Biomed. Health Inform. 2014, 20, 1404–1415.

27. Mellia, J.A.; Basta, M.N.; Toyoda, Y.; Othman, S.; Elfanagely, O.; Morris, M.P.; Torre-Healy, L.; Ungar, L.H.; Fischer, J.P. Natural Language Processing in Surgery: A Systematic Review and Meta-analysis. Ann. Surg. 2021, 273, 900–908.

28. Stahl, C.C.; Jung, S.A.; Rosser, A.A.; Kraut, A.S.; Schnapp, B.H.; Westergaard, M.; Hamedani, A.G.; Minter, R.M.; Greenberg, J.A. Natural language processing and entrustable professional activity text feedback in surgery: A machine learning model of resident autonomy. Am. J. Surg. 2021, 221, 369–375.

29. Gumbs, A.A.; Crovari, F.; Vidal, C.; Henri, P.; Gayet, B. Modified Robotic Lightweight Endoscope (ViKY) Validation In Vivo in a Porcine Model. Surg. Innov. 2007, 14, 261–264.

30. Gumbs, A.A.; Croner, R.; Rodriguez, A.; Zuker, N.; Perrakis, A.; Gayet, B. 200 Consecutive laparoscopic pancreatic resections performed with a robotically controlled laparoscope holder. Surg. Endosc. 2013, 27, 3781–3791.