The American Journal of Surgery 220 (2020) 1372–1374

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

My Thoughts / My Surgical Practice

Ethical thinking machines in surgery and the requirement for clinical leadership

Machine learning is making inroads into medicine and, despite a belief that procedural specialties may be more immune to this incursion, there is a growing sense that the cognitive functions of surgeons will at some point be automated by recursively self-improving artificial intelligence (AI) systems. AI systems are often criticized as biased, opaque, inscrutable or ethically ambiguous, yet a thorough self-reflection reveals that these same criticisms can be leveled at human decision-makers. Human bias occurs despite our best efforts and while it is challenging to gain insight into the decision-making processes of a deep neural network, the same difficulty holds true for clinicians, especially when some of this bias may be unconscious or deemed insignificant. A surgeon's explanation of a clinical decision when viewed through the filter of cognitive science is merely a credible, post-hoc rationalisation of the decision-making outcome, which may ignore the true nature of the partially unconscious, experience-driven, pattern-recognition-based algorithm that led to that decision. Surgeons, even those with substantial experience, are often inaccurate when making predictions regarding patient outcomes.[1] All humans are susceptible to the influence of heuristics, particularly when fatigue or stress set in. AI systems do not suffer from these corporeal limitations and their capabilities frequently exceed human prognostic accuracy with an unwavering cognitive capacity.[2]

We suggest that substantial quality and safety improvements can be gained from applying AI systems to surgery by leveraging their strengths and computational ability, and that robust clinician involvement is required to continually assess the potential negative effects of algorithm supported thinking while realizing that our human decision-making process is itself not perfect as demonstrated by the continual advancements and improvements in surgical medicine.[3] We present herein a high-level conceptual model to help facilitate AI system development and implementation, with the goal of achieving surgical quality and safety improvements.

Optimal outcomes in medicine are not just physiological but are also psychological. Eye contact, authenticity, and an ability to work in the context of unpredictable human behavior are capabilities that cannot yet be effectively addressed within narrow AI systems. Clinicians offer care but critically also offer the *perception* of care through the demonstration of empathy and ethical judgment. This perception is crucial especially when considered against the framework of the physician's oath to "First, do no harm." The concern for a patient's wellbeing and the physician's varied roles in that wellbeing may lead to a professional existential fear of automation, stemming partly from a concern that if algorithms replace physicians, and then malfunction, the contract between physician and patient may have been breached. An optimal clinical system will demand that the power of machines to process high-volume digitized complexity is linked with overt evidence of emotional intelligence and human judgement, thereby effectively leveraging the complementary strengths of human and machine intelligence.[4]

Predictions have been made regarding the full or partial automation of radiologists, pathologists, anesthesiologists and critical care specialists. Surgeons are not immune to this push towards automation and AI system implementation. Insofar as surgery is a systematic, data-driven, evidence-based, diagnostic and prognostic practice, machine learning has the potential to execute some component of these tasks with greater accuracy and consistency. This scenario may seem to be in the distant future given the present limitations of today's machine learning systems. A simplistic comparison of surgical capabilities between humans and machines may not yield the most relevant or beneficial strategies however; rather, considering the utility of surgical augmentation rather than automation may provide the most direct route to improve surgical outcomes.[4] AI is likely to exert substantial influence on many aspects of surgical practice.[3,5] High-performing machine learning models have been trained to predict adverse surgical outcomes, postoperative patient function and recovery, mortality, survival, and resource utilization through the processing and analysis of high volumes of medical record and radiomics data.[3] Machine learning systems can accurately predict mortality and adverse event risk in numerous surgical specialties.[3,6] In the field of neurosurgery, accurate non-invasive brain tumor classification has the potential to eliminate the need for entire surgical procedures along with their associated costs and surgical risks. Although surgical image-guidance systems have been in use within neurosurgery for well over three decades, AI is playing a central role in the development and optimization of robotic surgical systems that can deliver accuracy beyond that of standard human operator based systems. Surgeons may face a challenge acknowledging our individual limitations and being willing to include a faster and more accurate cognitive agent in the surgical team. Hopefully an understanding of medicine's growing complexity in the face of increasing healthcare resource scarcity will drive surgeons towards the improved efficiency and outcomes that combined systems may provide. It is clear that although sophisticated machine learning models that predict a

https://doi.org/10.1016/j.amjsurg.2020.06.073

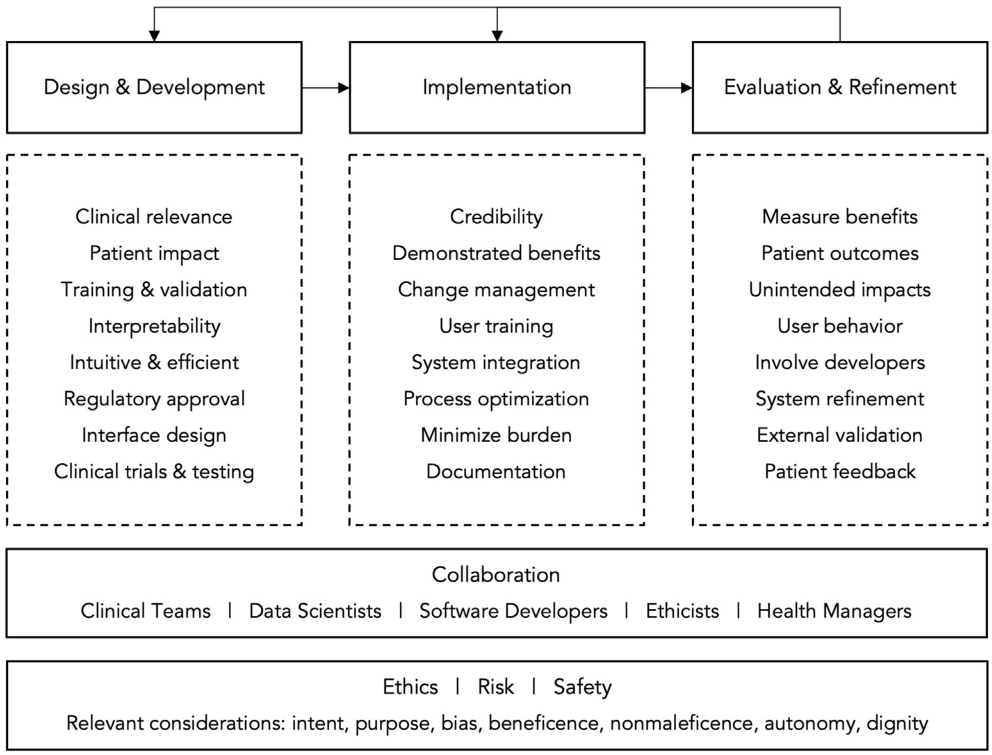

Fig. 1. A conceptual model to guide the development, implementation and evaluation of machine learning systems in surgical practice.

broad array of surgical outcomes are beneficial scientific advancements, their use may be ineffective or damaging if they are poorly designed and implemented without sufficient clinical insight. Often, systems are designed by engineers with little consideration of practical clinical issues or constraints. Clinical experience needs to be integrated into these systems and clinicians need to have a hand in their development. This clinical involvement will be critical to bridging the gap from systems that can predict salient or technical outcomes at the cost of intensive resources or impractical and expensive data inputs to those that can be realistically translated into clinical practice. This move towards useable clinical systems remains a challenge. There is a need for surgeons to learn about and understand these systems and their associated benefits and risks in order to have a capable hand in building them to deliver better care to patients. Just as it is difficult to imagine a clinician without a basic understanding of statistics in today's era of evidence-based medicine, so too must the surgeons of tomorrow have insight into data science and closer partnerships with data science practitioners.

Ethical and safety issues must be intentionally and appropriately addressed. Ethics must be embedded into clinical data science research from its inception and in the application of the research outcomes in practice. Systems must be designed with the best interests of patients and the community as their primary concern, adhering to established fundamental ethical principles.[7] It must be acknowledged that building human bias into trained algorithms is inevitable. Appropriate and effective mitigation of the subtle biases inherent in all training datasets and modeling methods must be carried out. A major present concern is that biased algorithms will widen existing health discrepancies. Effective model design and development requires adherence to guidelines that inform rigorous and appropriate algorithm selection, training, calibration and validation methods.[8] Clear and overt goal-setting and ethical framing may help to identify and mitigate biases in the system design phase. Establishing clear ethical goals will be critical, as the current example of AI in advancements of self-driving cars demonstrates. Today's technology is likely to get a vehicle from point A to point B consistently but cannot yet assure avoidance of harm to human life. Delineating the primary goals of each, such a system will help guide a route towards success. Post-implementation evaluation activities will need to be vigilant for the unintended effects of predictive clinical decision support systems. Privacy and consent are critical concerns when considering the use of patient data for algorithm training and validation. In short, ethical principles need to be core considerations at all stages, from system design and development, through to implementation, evaluation and refinement.

A high-level conceptual model is presented (Fig. 1) to facilitate the implementation of AI systems to surgical practice, leveraging established critical success and failure factors.[4] The model consists of three operational components laid out sequentially: (1) initial design and development, (2) implementation, and (3) evaluation and refinement. Each component contains a set of core considerations that are likely to drive success. Feedback loops link the third component (evaluation and refinement) with the first two components, indicating the necessity of continuous improvement activities. Of particular note, the model is underpinned by the bedrock of effective multidisciplinary collaboration, ethics, safety and risk management. This nascent model is based on the experience of numerous domain experts and a survey of the literature. It is presented as a practical starting point for clinical leaders and is currently undergoing further development and validation.

If clearly-stated objectives and control of AI system objective optimization are key to safe and successful implementation,[9] then clinicians need to begin learning and leading now. Numerous detailed primers exist to facilitate this process.[4] To realise the substantial potential benefits of AI systems in surgery, we need to move from proof of concept to informed implementation. This transition will require clinical leadership with effective and ethical development and evaluation capabilities.[4] The creation of beneficial machine-assisted decision-making systems requires close attention to ethical pitfalls and unintended consequences at all stages of development and implementation. As we draw incrementally closer to AI systems capable of recursive self-improvement, now is the time for surgeons to prepare for a more automated future that aligns with the best interests of patients.

Declaration of competing interest

All authors have reviewed and approved this manuscript and have no relevant financial or other conflicts of interest with regard to this research and its publication.

Acknowledgement

The research team would like to thank Professor Margaret Somerville for sharing her insights.

References

1. Katlic MR, Coleman J. Surgical intuition. *Ann Surg*. 2018;268:935–937.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44–56.
3. Buchlak QD, Esmaili N, Leveque J-C, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. *Neurosurg Rev*. 2019;1–19. https://doi.org/10.1007/s10143-019-01163-8.
4. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25:30–36.
5. Kusminsky RE. The physician of the future and the future of physicians. *Am J Surg*. 2019;217:811–812.
6. Loftus TJ, Tighe PJ, Filiberto AC, et al. Opportunities for machine learning to improve surgical ward safety. *Am J Surg*. 2020. Available online 26 February 2020.
7. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med*. 2018;378:981.
8. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res*. 2016;18:e323.
9. Russell S, Dewey D, Tegmark M. Research priorities for robust and beneficial artificial intelligence. *AI Mag*. 2015;36:105–114.

Quinlan D. Buchlak*
School of Medicine, The University of Notre Dame Australia, Sydney, NSW, Australia

Nazanin Esmaili
School of Medicine, The University of Notre Dame Australia, Sydney, NSW, Australia

Department of Medicine, University of Toronto, Toronto, ON, Canada

Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW, Australia

Jean-Christophe Leveque
Neuroscience Institute, Virginia Mason Medical Center, Seattle, WA, USA

Christine Bennett
School of Medicine, The University of Notre Dame Australia, Sydney, NSW, Australia

Massimo Piccardi
Faculty of Engineering and IT, University of Technology Sydney, Ultimo, NSW, Australia

Farrokh Farrokhi
Neuroscience Institute, Virginia Mason Medical Center, Seattle, WA, USA

* Corresponding author. The University of Notre Dame Australia, 160 Oxford St, Sydney, 2015, Australia.
E-mail address: quinlan.buchlak1@my.nd.edu.au (Q.D. Buchlak).

26 June 2020