Viewpoint

Tempering Expectations on the Medical Artificial Intelligence Revolution: The Medical Trainee Viewpoint

Zoe Hu[1*], BSc, MD; Ricky Hu[1,2*], BASc, MASc; Olivia Yau[3], BSc; Minnie Teng[3], MSc; Patrick Wang[1], BHSc, MD; Grace Hu[4], BSc; Rohit Singla[2,3], BASc, MASc

[1]School of Medicine, Queen's University, Kingston, ON, Canada

[2]School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada

[3]School of Medicine, University of British Columbia, Vancouver, BC, Canada

[4]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Zoe Hu, BSc, MD
School of Medicine
Queen's University
166 Brock Street
Kingston, ON, K7L5G2
Canada
Phone: 1 6132042952
Email: zhu@qmed.ca

Abstract

The rapid development of artificial intelligence (AI) in medicine has resulted in an increased number of applications deployed in clinical trials. AI tools have been developed with goals of improving diagnostic accuracy, workflow efficiency through automation, and discovery of novel features in clinical data. There is subsequent concern on the role of AI in replacing existing tasks traditionally entrusted to physicians. This has implications for medical trainees who may make decisions based on the perception of how disruptive AI may be to their future career. This commentary discusses current barriers to AI adoption to moderate concerns of the role of AI in the clinical setting, particularly as a standalone tool that replaces physicians. Technical limitations of AI include generalizability of performance and deficits in existing infrastructure to accommodate data, both of which are less obvious in pilot studies, where high performance is achieved in a controlled data processing environment. Economic limitations include rigorous regulatory requirements to deploy medical devices safely, particularly if AI is to replace human decision-making. Ethical guidelines are also required in the event of dysfunction to identify responsibility of the developer of the tool, health care authority, and patient. The consequences are apparent when identifying the scope of existing AI tools, most of which aim to be physician assisting rather than a physician replacement. The combination of the limitations will delay the onset of ubiquitous AI tools that perform standalone clinical tasks. The role of the physician likely remains paramount to clinical decision-making in the near future.

Introduction

The field of artificial intelligence (AI) in medicine has seen rapid development in the last decade, with an increasing number of applications introduced in clinical settings [1]. With the rapid growth in computing power and data, medical AI has transformed from an afterthought into an imminent possibility.

Currently, the utility of AI in completing tasks such as diagnostic prediction, automation, and generation of features from clinical data is recognized in many specialties. Models predicted the incidence of myocardial infarction and outperformed the current gold standard American College of Cardiology and American Heart Association risk algorithm [2]. These technological advancements have understandably raised concerns among health care trainees and professionals that AI may be taking over their duties. A study assessing medical students' views

regarding the impact of AI on future careers reported that 78.77% (1707/2167) expect significant changes due to AI and 89.62% (1942/2167) expressed that careful supervision by humans is required [3].

To moderate the concerns of AI in disrupting the future role of physicians, an understanding of the capabilities and limitations of AI tools is required. Wiens et al [4] reported AI adoption challenges, including problem formulation to market transition, all of which will require cooperation with interdisciplinary teams and systemwide change. In addition to refining the results of an AI algorithm, how the results are conveyed must also be accepted. Even if a physician accepts the judgement of a computer as legitimate, patients may not be nearly as receptive.

The aim of this commentary is to analyze the multifaceted issue of medical AI adoption to temper preconceived notions regarding its impact and rapid progression. We identify and explore four major barriers to AI adoption: (1) the limitations of performance and biases in AI applications, (2) the limitations due to heterogeneous digital infrastructure, (3) the limitations due to lack of technological literacy, and (4) the limitations of ethical challenges associated with medical AI usage.

Limitations of Performance

A significant barrier for AI applications to be implemented is regulatory approval, such as by the Food and Drug Administration (FDA), where AI applications would be included in the recently created category of Software as a Medical Device [5]. Certification is required for a recognized regulatory body to approve of a device's safety and effectiveness. If a new medical device is not considered a low- or moderate-risk device, it is required to enter the stringent premarket approval pathway, where demonstration of safety and effectiveness is required from clinical studies. The device is also classified in risk classes from Class I (the lowest risk) to Class III (the highest risk) [5]. AI, particularly machine learning, poses unique challenges as a machine learning model may continuously update with new training data. As such, the FDA has created recent guidelines, indicating that surveillance is required over the total product life cycle of the device, including model updates from retraining [6].

A standalone diagnostic tool would likely enter the premarket approval pathway and require extensive testing such as randomized controlled trials [7]. Leeuwen et al [8] evaluated 100 AI devices with CE-marked approval in Europe and reported that only 2 products were classified as class III, requiring premarket approval. Of 100 AI devices, 64 had no peer-reviewed studies validating the product performance. Wu et al [9] evaluated 54 AI medical devices approved by the FDA, with none being standalone diagnostic devices without physician supervision and none tested in a prospective trial. Hence, the current state of AI devices toward the FDA label of Computer-Assisted Detection Devices, which pose less resistance for market entry. The financial incentive results in a trend of devices being developed as physician-assisting tools that physicians can use at their discretion [10].

A technical barrier for AI devices to replace human analysis is the current performance of AI devices. For instance, when validated on a data set from a single center, convolutional neural networks (CNNs) routinely achieve accuracies above 0.90 [11]. However, with the variability of medical imaging from different machines, operators, or imaging protocols, multicenter studies are required to validate the generalizability of these classifiers. Alice et al [11] reported that 81% of diagnostic algorithms reviewed results in significant decrease of accuracy when externally validated. Thus, rigorous validation is required with a diverse data set to address the major machine learning challenges of data scarcity, population shifts from different data sets, prevalence shifts, and selection biases [12]. External validation also reveals a more accurate comparison between human and machine performance. Rodriguez-Ruiz et al [13] reported that when testing a published CNN to classify malignancies from mammography on a data set of 2652 images from seven different countries, the CNN performed within the same 95% CI accuracy range of 101 different radiologists [13].

The rigorous validation requirements for AI to be usable in clinical practice is evident when analyzing rapidly developed AI models. In the COVID-19 pandemic, over 100 diagnostic prediction models have been trained and published in literature, using features such as chest x-ray data, lung ultrasound, vital signs, and lab values. The reported concordance index of such models ranged from 0.71-0.99. However, Wynants et al [14] assessed that only 5% of the models found performed external validation, and only 2 models addressed selection biases during sampling.

An additional challenge for AI applications is that the ability to learn complex features is restricted to the architecture of the AI model. For instance, medical applications for CNNs commonly use architectures that perform well on the ImageNet challenge. The CNN architecture defines model parameters such as resolution, depth, and number of input channels, all of which affect the ability to detect complex features related to some objective. However, newer architectures are frequently developed, such as EfficientNet outperforming ResNet, DenseNet, Xception, and ResNeXT, all of which have been previously used in medical image classifiers [15]. Updating the model architecture is a significant change to the model. For instance, ResNet introduces the usage of residual blocks in a layer as an input for a subsequent layer to begin learning, changing how the model is initialized. This may require reapproval from regulatory bodies due to nontrivial changes in the device.

The alternative of a physician-assisting device is more likely in the near future, such as automating report extraction from imaging studies or image reconstruction to reduce excessive radiation from repeated imaging [16,17]. This reduces competition with physician tasks while still providing clinical utility from complex AI analyses.

Limitations of Current Infrastructure

Implementation of an AI product, even with validated performance, is limited by heterogenous digital infrastructure in health care systems. Different areas of patient care such as

inpatient progress notes, laboratory results, and discharge summaries may all have independent databases. This complexity is further multiplied by interactions with outpatient clinics and health authorities across provincial or state boundaries.

The incomplete adoption of electronic medical records (EMRs) illustrates the lag in digital infrastructure integration despite electronic record technology being available. The Canadian Federal Government's Economic Action Plan provided funding to health care providers toward establishing EMRs in primary care in 2010, leading to an increase of EMR adoption [18]. A similar progression took place in the United States in 2014 [19]. Despite this, there continues to be reliance on paper files in both primary care clinics and hospitals [20]. If, for instance, an algorithm in an emergency department requires baseline laboratory markers for a patient from their family physician, then standardization and likely digitalization of the input data is required.

There are currently 11 certified EMR vendors and 12 EMR products in Ontario [21]. Although hospitals often have a primary vendor, they often employ a variety of disparate EMR products in affiliated practices [21]. In theory, digitization of health care data would provide an abundance of high-quality data for AI research. However, EMR vendors operate in silos and use their own approach to storing data. To implement an AI product in practice may necessitate creation of a completely novel data pipeline to aggregate records across different databases. There are attempts at standardization including the "EMR Content Standard" by the Canadian Institute for Health Informatics [22]. This introduces a content standard for EMR data entry, but levels of prioritization of the standard differ across provinces, and no standard EMR data entry has been universally adopted, resulting in the persistence of difficulty in coalescing data to be usable by AI.

For AI technology to be successful, patients must consent to its use and trust the safety of the technology. A recent public opinion survey in the United States on AI indicated that data privacy was considered to be the most important issue [23]. Privacy concerns and restricted access limits access to a diverse and large sample size, which is necessary for an AI algorithm to be validated and implemented in clinical practice [24]. A diverse data set is also crucial to guarantee adequate representation of patient cohorts in AI algorithm training [25]. There are approaches to overcome these barriers including federated learning, where a model is shared across different centers for training without exporting data [24]. However, these approaches require universal agreements regarding scope and are currently not standard of practice.

Limitations of Technological Literacy

Medical AI applications have become increasingly relevant at an accelerated rate, though the lag in technological literacy of health care professionals for AI technology exceeds the expected social and cognitive lag of adapting new technology [26]. One challenge is that there is currently no standardized curriculum for AI education nor are there any relevant accreditation requirements within most medical doctorate programs [27]. This gap is significant as health care professionals are the main users of medical AI applications and will have to be responsible for appropriate usage of AI applications [28].

Despite a recent surge in interest in training health care trainees in AI, universal integration of AI education into current health care training is a nontrivial challenge. Medical training is dense and rigorous with significant demands on trainees and staff [29]. Implementation of such a curriculum also requires specific faculty expertise. Even with qualified educators available, there is the challenge of selecting the correct depth and breadth of topics required for medical trainees.

Without appropriate medical AI education, health care professionals may not be adequately equipped to navigate the potential ethical and legal implications of AI in health care. The flexibility that health care providers have in using their judgement to make clinical decisions tailored to an individual patient, using contextual understanding of interpatient and intrapatient variations, is essential to medicine. This process may be impeded if the end user lacks the basic digital literacy to understand the limitations of such applications of AI; for instance, deciding when to override an AI analysis in favor of contextual clinical judgement or vice versa. However, acquiring digital competency in AI applications may imply time away from service for health care providers and extra study workload for health care trainees, in addition to growing medical knowledge. Other challenges that contribute to the gap in technological literacy include lack of awareness of digital knowledge required for health care, lack of equitable access to AI education, and limited trust in AI applications in health care.

Medical applications must be well performing, trustworthy, transparent, interpretable, and explainable. Interpretation of AI models requires technical training, making it difficult to assess its performance. This is especially true in complex AI models such as deep neural networks, where it is not often possible to examine what features are used to compute the output, creating a colloquial "black box" algorithm. The gap in technological literacy among health care professionals, which is further hindered by the difficulty in implementing AI literacy training of an appropriate scope, prevents many AI applications from advancing beyond the proof-of-concept "computer-side" stage to bedside application [30].

Limitations of Ethical Challenges

In the presence of errors by AI decisions, there lies challenges not only in identifying liability but also in quality improvement analysis. Harm caused by AI may be due to several reasons in the pipeline, such as poor data stewardship, incomplete mathematical constraints resulting in an inaccurate model, or inappropriate usage by a clinician [31]. For instance, if an AI algorithm misdiagnoses a patient, causing an adverse event, is the error associated with data collection that was not representative of patient characteristics, with inadequate algorithm development resulting in computations that produce an inaccurate prediction, or with health care administration for deciding to use an AI product? Traditional quality improvement analysis in medicine, such as cause-effect analysis, may be insufficient because it lacks a 1-dimensional cause-to-effect pathway, particularly with multiparametric AI models such as

neural networks, which contain millions of computational kernels [32]. Interdisciplinary collaboration between data scientists, data stewards, clinicians, and health care workers is crucial to developing a risk liability and quality improvement system before AI can serve as a medical decision maker.

Additionally, substantial data bias may lead to unforeseen disparities in patient care as AI may stratify based on unintentional subgroups. Gichoya et al [33] observed that chest x-ray AI models can be used to predict patient's race with image features physicians were unaware of. The implication is that bias is unavoidable even when looking at data that appears agnostic, such as chest x-rays. This may further encourage health care disparities if the model makes decisions directly correlated with race or gender. There is then a utilitarian conflict of beneficence in deciding the extent to which it is acceptable to use an AI algorithm that may be more accurate and benefit certain subgroups at the expense of others; for instance, triaging resources for subgroups that AI can accurately analyze. There is also a deontological conflict to adhere to nonmaleficence. If

we know there is a high likelihood of increasing disparity despite the beneficial aspects of AI, the application of AI would be unethical.

Hence, AI poses unique ethical issues due to limitations of transparency and inherent potential for harm when used as a decision maker. AI is capable of identifying hidden features within data that can be leveraged to improve decision-making, but it is not without potential risk and needs to be deliberated by all stakeholders involved in the process.

Conclusions

Implementation of AI in medicine faces barriers of regulatory approval, performance, compatibility of digital infrastructure, and shared multidisciplinary collaboration. Although AI shows potential in improving quality of life for patients by enhancing decision-making and tasks carried by health care professionals, the adoption of AI is likely incremental rather than a stark change in standard of care.

Conflicts of Interest

None declared.

References

1. Briganti G, Le Moine O. Artificial intelligence in medicine: today and tomorrow. Front Med (Lausanne) 2020 Feb 5;7:27 [FREE Full text] [doi: 10.3389/fmed.2020.00027] [Medline: 32118012]
2. Deo RC. Machine learning in medicine. Circulation 2015 Nov 17;132(20):1920-1930 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.001593] [Medline: 26572668]
3. Teng M, Singla R, Yau O, Lamoureux D, Gupta A, Hu Z, et al. Health care students' perspectives on artificial intelligence: countrywide survey in Canada. JMIR Med Educ 2022 Jan 31;8(1):e33390 [FREE Full text] [doi: 10.2196/33390] [Medline: 35099397]
4. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019 Sep 19;25(9):1337-1340. [doi: 10.1038/s41591-019-0548-6] [Medline: 31427808]
5. Smith JA, Abhari RE, Hussain Z, Heneghan C, Collins GS, Carr AJ. Industry ties and evidence in public comments on the FDA framework for modifications to artificial intelligence/machine learning-based medical devices: a cross sectional study. BMJ Open 2020 Oct 14;10(10):e039969 [FREE Full text] [doi: 10.1136/bmjopen-2020-039969] [Medline: 33055121]
6. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 2020 Sep 11;3(1):118 [FREE Full text] [doi: 10.1038/s41746-020-00324-0] [Medline: 32984550]
7. Harvey HB, Gowda V. How the FDA regulates AI. Acad Radiol 2020 Jan;27(1):58-61. [doi: 10.1016/j.acra.2019.09.017] [Medline: 31818387]
8. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol 2021 Jun 15;31(6):3797-3804 [FREE Full text] [doi: 10.1007/s00330-021-07892-z] [Medline: 33856519]
9. Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nat Med 2021 Apr;27(4):582-584. [doi: 10.1038/s41591-021-01312-x] [Medline: 33820998]
10. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ Digit Med 2021 Apr 07;4(1):65 [FREE Full text] [doi: 10.1038/s41746-021-00438-z] [Medline: 33828217]
11. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: a systematic review. Radiol Artif Intell 2022 May 01;4(3):e210064 [FREE Full text] [doi: 10.1148/ryai.210064] [Medline: 35652114]
12. Castro DC, Walker I, Glocker B. Causality matters in medical imaging. Nat Commun 2020 Jul 22;11(1):3673 [FREE Full text] [doi: 10.1038/s41467-020-17478-w] [Medline: 32699250]
13. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. J Natl Cancer Inst 2019 Mar 05:916-922. [doi: 10.1093/jnci/djy222] [Medline: 30834436]

14. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]

15. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. Preprint posted online on May 28, 2019 . [doi: 10.48550/arXiv.1905.11946]

16. Carrodeguas E, Lacson R, Swanson W, Khorasani R. Use of machine learning to identify follow-up recommendations in radiology reports. J Am Coll Radiol 2019 Mar;16(3):336-343 [FREE Full text] [doi: 10.1016/j.jacr.2018.10.020] [Medline: 30600162]

17. Kambadakone A. Artificial intelligence and CT image reconstruction: potential of a new era in radiation dose reduction. J Am Coll Radiol 2020 May;17(5):649-651. [doi: 10.1016/j.jacr.2019.12.025] [Medline: 32004484]

18. Zhao EJ. The future of electronic medical records in Canada. CMAJ 2019 May 13;191(19):E542-E542 [FREE Full text] [doi: 10.1503/cmaj.71858] [Medline: 31085568]

19. Bristol N. The muddle of US electronic medical records. The Lancet 2005 May;365(9471):1610-1611. [doi: 10.1016/s0140-6736(05)66492-6]

20. Saleem JJ, Russ AL, Justice CF, Hagg H, Ebright PR, Woodbridge PA, et al. Exploring the persistence of paper with the electronic health record. Int J Med Inform 2009 Sep;78(9):618-628. [doi: 10.1016/j.ijmedinf.2009.04.001] [Medline: 19464231]

21. Larsen D, Hutchison S. Single electronic medical record for Canada: a second opinion. CMAJ 2019 May 13;191(19):E539-E540 [FREE Full text] [doi: 10.1503/cmaj.71810] [Medline: 31085566]

22. Keshavjee K, Williamson T, Martin K, Truant R, Aliarzadeh B, Ghany A, et al. Getting to usable EMR data. Can Fam Physician 2014 Apr;60(4):392 [FREE Full text] [Medline: 24733333]

23. Singh RP, Hom GL, Abramoff MD, Campbell JP, Chiang MF, AAO Task Force on Artificial Intelligence. Current challenges and barriers to real-world artificial intelligence adoption for the healthcare system, provider, and the patient. Transl Vis Sci Technol 2020 Aug 28;9(2):45 [FREE Full text] [doi: 10.1167/tvst.9.2.45] [Medline: 32879755]

24. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med 2020 Sep 14;3(1):119 [FREE Full text] [doi: 10.1038/s41746-020-00323-1] [Medline: 33015372]

25. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med 2019;2:77 [FREE Full text] [doi: 10.1038/s41746-019-0155-4] [Medline: 31453372]

26. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med 2019;2:79 [FREE Full text] [doi: 10.1038/s41746-019-0158-1] [Medline: 31453374]

27. Kolachalama VB, Garg PS. Machine learning and medical education. NPJ Digit Med 2018 Sep 27;1(1):54 [FREE Full text] [doi: 10.1038/s41746-018-0061-1] [Medline: 31304333]

28. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan 7;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]

29. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. J Intern Med 2018 Jun 24;283(6):516-529 [FREE Full text] [doi: 10.1111/joim.12752] [Medline: 29505159]

30. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195 [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]

31. Smith H, Fotheringham K. Artificial intelligence in clinical decision-making: rethinking liability. Med Law Int 2020 Aug 26;20(2):131-154. [doi: 10.1177/0968533220945766]

32. Plsek PE. Quality improvement methods in clinical medicine. Pediatrics 1999 Jan;103(1 Suppl E):203-214. [Medline: 9917464]

33. Gichoya J, Banerjee I, Bhimireddy A, Burns J, Celi L, Chen L, et al. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health 2022 Jun;4(6):e406-e414 [FREE Full text] [doi: 10.1016/S2589-7500(22)00063-2] [Medline: 35568690]

Abbreviations

AI: artificial intelligence
CNN: convolutional neural network
EMR: electronic medical record
FDA: Food and Drug Administration

XSL•FO
RenderX