Palmer et al. Critical Care (2019) 23:136
https://doi.org/10.1186/s13054-019-2424-7

Critical Care

EDITORIAL Open Access

Intelligently learning from data

Edward Palmer[1*], Roman Klapaukh[2], Steve Harris[1,3], Mervyn Singer[1,3] and the INFORM-lab

Keywords: Artificial intelligence, Machine learning, Statistical models

Main text

Methods from the fields of artificial intelligence (AI) and machine learning (ML) are entering the medical literature at an unprecedented rate. A PubMed search using the keyword "Machine Learning" has shown an accelerating year-on-year increase in publications. Leveraged with "big data", these approaches are often lauded as transformative in healthcare with the promise that they can and will solve all of our problems [1]. While these developments are indeed exciting, we caution the need to place realistic constraints on our expectations.

There is an established history of computational learning by fitting models to data. Previously the purview of statisticians, these models help us to understand complex problems by identifying patterns in data that are otherwise unnoticeable to humans. These models are typically associative in design, in that correlations do not necessarily imply causation. Given the well-described limitations of statistical models, there is a healthy scepticism of these approaches. Despite an awareness of these limitations, humans seem hard-wired to see a causal paradigm in mathematical models [2].

AI and ML models are a set of methods for learning patterns from data. Albeit optimised for different scenarios, statistical and ML approaches both share the same goal. AI models emphasise predictive accuracy, typically in large datasets, without a particular focus on inference for any one individual predictor. Statistical models stress a direct analytical approach, which characterises with uncertainty, estimators for individual predictors [3]. Statistical models tend to provide parameters that have a more directly interpretable human meaning. This ease of interpretability can make statistical models less able to describe complex phenomena. They are often either intractable at scale, or become powered to detect clinically meaningless signals. With these shortcomings in mind, AI models have enabled a new branch of learning from massive datasets.

AI models excel where they train on large volumes of high-quality labelled data. The prototypical example of which in medicine is computed tomography disease detection [4]. In these scenarios, predictive accuracy is the primary goal, and causal inferences are not necessary. Other examples of AI with direct relevance to this journal's readership include early detection of the deteriorating patient [5] and strategies for fluid and inotropic administration in sepsis [6]. Not being conducted under a causal framework is a vital caveat to such approaches [7], yet our experience reveals a tendency for clinicians to draw causal conclusions from these models. While causal AI is a burgeoning field, generally AI models do not address the fundamental problem of causal inference, i.e. that one can only observe a single outcome for each patient (did the patient live or die), and not the counterfactual outcomes (how would the patient have responded under different treatments). Fundamentally, a model cannot learn from data that does not exist.

Judea Pearl, a professor of computer science and causal inference, describes AI models as running "almost entirely in an associational mode" [8]. Many of these approaches do not model the flow of time (that effect must follow cause) as associative models are true in either direction.

Spurious associations are commonplace when learning from data. These range from trivial (e.g. Facebook likes for "curly fries" are highly predictive of a high IQ [9]) to deeply concerning (e.g. asthma is predictive for a good outcome in pneumonia) [10]. AI models can be highly sensitive to the data on which they were trained, the addition of imperceptible noise, and improperly defined intermediate rewards (the means through which reinforcement models learn associations). This approach can lead to unexpected results, without a clear explanation of how or why this occurred [11, 12]. Komorowski

* Correspondence: edward.palmer@ucl.ac.uk
[1]Bloomsbury Institute of Intensive Care Medicine, University College London, London, UK
Full list of author information is available at the end of the article

et al. identify an optimal treatment strategy for sepsis that suggested less fluid administration than the human clinician. It is difficult to discern if this finding is causal (reducing fluid administration in septic shock will improve outcomes) or associational (better outcomes are seen for those patients who require less fluid). If these models are applied indiscriminately and without application of strong domain knowledge, spurious inferences are often found. Patients could potentially come to harm by extrapolating such findings to the bedside without a rigorous understanding of the causal pathways or underlying mechanisms that are typically discovered through experimental research.

Trusting and implementing AI models prematurely, just because they are "new" and therefore perceived as "better", could lead to a lack of trust in these critical methods. The recent Topol review for the National Health Service [13] emphasised the importance of strengthening links between clinical practice and data science. As AI models enter the literature, and the medical community considers incorporating them into clinical decision-making tools, it is imperative that any outputs are both understood and challenged constructively. The rate at which new methods are appearing in the AI literature can leave little time to examine their limitations in a clinical context; impressive technical progress may outstrip the pace at which it is safe to implement.

Models of the real world, regardless of their origin, are still models. Models allow us to understand complex processes and, when safe and proper to do so, take actions based on their recommendations. AI models are compelling and provide rich insights into the world in which we practise. However, following either an AI model or statistical model without due care and consideration could place patients at risk.

Abbreviations
AI: Artificial intelligence; IQ: Intelligence quotient; ML: Machine learning

Acknowledgements
The INFORM-lab: Tim Bonnici, Ahmed Al-Hindawi, Tom Keen.

Funding
None.

Availability of data and materials
Not applicable.

Authors' contributions
All authors contributed equally to the writing of this editorial. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Bloomsbury Institute of Intensive Care Medicine, University College London, London, UK. [2]Research Software Development Group, Research IT Services, University College London, London, UK. [3]Critical Care, University College London Hospitals NHS Foundation Trust, London, UK.

Received: 25 March 2019 Accepted: 8 April 2019
Published online: 24 April 2019

References
1. The future of healthcare: our vision for digital, data and technology in health and care. Policy paper. https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care. Accessed 11 Mar 2019
2. Bordacconi MJ, Larsen MV. Regression to causality: regression-style presentation influences causal attribution. Res Polit. 2014;1(2):1–5.
3. Harrell F. Road map for choosing between statistical modeling and machine learning. http://www.fharrell.com/post/stat-ml/.. Accessed 20 Mar 2019
4. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp. 2018;2(1):35.
5. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An interpretable machine learning model for accurate prediction of sepsis in the ICU. Crit Care Med. 2018;46(4):547–53.
6. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med. 2018;24(11):1716–20.
7. Jeter R, Josef C, Shashikumar S, Nemati S. Does the "Artificial Intelligence Clinician" learn optimal treatment strategies for sepsis in intensive care? arXiv:190203271. http://arxiv.org/abs/1902.03271. Accessed 16 Mar 2019
8. Pearl J, Mackenzie D, Penguin. The book of why: the new science of cause and effect. London: Allen Lane; 2018.
9. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci. 2013;110(15):5802–5.
10. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining '15. Sydney: ACM Press; 2015. p. 1721–30.
11. Gottesman O, Johansson F, Meier J, Dent J, Lee D, Srinivasan S, et al. Evaluating reinforcement learning algorithms in observational health settings. arXiv:180512298. http://arxiv.org/abs/1805.12298. Accessed 11 Mar 2019.
12. Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mané D. Concrete problems in AI safety. arXiv:160606565. http://arxiv.org/abs/1606.06565. Accessed 13 Mar 2019
13. Topol Review. Health Education England policy document. https://www.hee.nhs.uk/our-work/topol-review. Accessed 20 Mar 2019