Bridging The Chasm Between Deep Learning and Clinical Implementation

In recent years, there have been advances in artificial intelligence (AI) for health care using deep neural networks, many of which have been commercialised. However, few AI tools have been implemented in health systems. Why has this chasm occurred?

The deployment of any new technology is usually managed centrally in hospitals and health systems. For the information technology (IT) teams, there is the concern that input data are drawn from outside the hospitals and health systems where the model aims to be implemented and algorithm performance, source code, and input data, are unavailable to review . Many commercial AI applications are in the radiology, but few are supported by evidence from published studies. Added to this lack of transparency is the concern that the algorithms were tested and validated using retrospective, in-silico data, which may not reflect real-world clinical practice. Regulators reviewing a company's AI data are privy to considerable data but these data are usually unavailable to health system IT or the medical community.

Transparency, suitability, and adaptability are not the only concerns. AI imported from a commercial setting into a hospital or health system can raise issues of equity, safety, necessary regulation not only because little is known about which data have been used to train the AI, but also because there are concerns about the ownership of the subsequent, real-patient data that AI draws on post-implementation. Additionally, AI implementations can be hindered by inadequate information about the data used to make decisions and recommendations. Clinicians receiving AI-generated decisions rarely have oversight of the datapoints used to reach a specific decision, contributing to so-called algorithmic aversion. Such algorithmic aversion may also arise from the uncertainty surrounding who (developer companies, researchers, clinicians, or hospitals) should accept responsibility for the algorithm's decisions. Any consideration by hospital leadership to implement AI-based diagnostic and decision-making tools should closely consider these concerns and barriers.

How can we bridge this chasm? Algorithmic aversion can be alleviated when the recipients of AI decisions can modify how the algorithm works or if they believe that algorithm can learn from local data. A three-fold approach can help optimise clinical use of AI. First, transparency into the datasets used for initial training of the AI tools, which would generate full trust for clinicians and patients. Second, the deconstruction of the neural networks to make the features that drive the AI performance understandable for clinicians. Third, allowing for clinicians to retrain AI models with local data if the needs of their populations and hospital requires it. Progress is likely to come with the development of open source AI trained on open data depositories and publicly shared algorithms. One example that shows the possible potential of this approach is Innereye, a clinically implemented AI for radiotherapy that has been implemented within a UK hospital and trained with hospital data and by hospital clinicians. If coupled with privacy-preserving computing tools such as federated learning, open-source AI could further remove barriers for the fast scalability of home-grown AI solutions developed in hospitals across the health system while maintaining clinician and patient trust in the ownership and regulation of data. Home-grown AI solutions, such as Mayo clinic's AI-assisted screening tool for advanced review of the electrocardiogram, with determination of cardiac function and potential missed clinical diagnoses, and the rapid, , AI-enabled genomics tool developed at Rady Children's Hospital for sick neonates without a diagnosis, point to how different prebuilt algorithmic tools and clinicians' expertise can come together to generate adapted AI to help cater to the specific needs of hospitals and patient populations. By mandating high standards for the implementation of open source, home-grown, and adapted AI tools, their implementation could be rigorous yet flexible enough to fit within a hospital's unique IT infrastructure and governance frameworks.

Ultimately, it will be necessary for clinicians to embrace AI tools for substantial implementation to occur. That acceptance will be potentiated by rigorous clinical validation, a high bar not reflected by what companies are typically willing to pursue. Currently, algorithms are not further refined and adapted after regulators provide approval, which limits the autodidactic feature of deep learning that could be a persuasive tool for the value of an algorithm for IT staff and clinicians. The more that is done to address these multiple barriers, the sooner we'll be able to see wider implementation of AI for routine patient care.

Angela Aristidou, Rajesh Jena, and Eric J. Topol

Angela Arisitdou University College London School of Management
a.aristidou@ucl.ac.uk

Rajesh Jena University of Cambridge, rajesh.jena@addenbrookes.nhs.uk

Eric Topol, Scrips Research, etopol@scripps.edu

Conflicts of interest:

Further Reading

Iacobucci G. Digital health: GPs aren't "Luddites" but want safe, equitable care. *BMJ* 2019; **364**: 1258.
Allen B, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 American College of Radiology Data Science Institute Artificial Intelligence Survey. *Journal of the American College of Radiology* 2021; **18**, 8: 1153-59.
Watson D.S, Krutzinna J, Bruce I.N, Griffiths C, McInnes I, Barnes M, Floridi L. Clinical applications of machine learning algorithms: beyond the black box. *BMJ* 2019; **364**: l886.
Longoni C, Bonezzi A, Morewedge C. Resistance to Medical Artificial Intelligence. *Journal of Consumer Research* 2019; **46**, 4: 629–650.
Dietvorst B, Simmons J, Massey C. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science* 2018; **64**, 3: 1155–70.
Berger B, Adam M, Ru A, Benlian A. Watch Me Improve — Algorithm Aversion and Demonstrating the Ability to Learn. *Business and Information Engineering* 2021; **63**, 1: 55–68.