The challenge of clinical adoption—the insurmountable obstacle that will stop machine learning?

1JONATHAN TAYLOR, MEng, MSc and 2JOHN FENNER, PhD

1Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
2University of Sheffield, Sheffield, UK

Address correspondence to: Mr Jonathan Taylor
E-mail: jonathan.taylor@sth.nhs.uk

ABSTRACT

Machine learning promises much in the field of radiology, both in terms of software that can directly analyse patient data and algorithms that can automatically perform other processes in the reporting pipeline. However, clinical practice remains largely untouched by such technology. This article highlights what we consider to be the major obstacles to widespread clinical adoption of machine learning software, namely: representative data and evidence, regulations, health economics, heterogeneity of the clinical environment and support and promotion. We argue that these issues are currently so substantial that machine learning will struggle to find acceptance beyond the narrow group of applications where the potential benefits are readily evident. In order that machine learning can fulfil its potential in radiology, a radical new approach is needed, where significant resources are directed at reducing impediments to translation rather than always being focused solely on development of the technology itself.

Machine learning is a topic of major interest in several areas of medicine. With a history of success in non-medical image analysis problems, it promises disruptive and transformative change in radiology.1 In particular, it has been developed for computer-aided diagnosis and detection applications, and for data processing tasks such as automated tumour volume measurements. However, despite the media attention and impressive results generated in the lab, where, for instance, machines have been shown to outperform radiologists in specific disease recognition tasks,2 uptake in the clinic remains vanishingly small. In the following sections, we highlight some of the substantial challenges of clinical translation which we believe have so far blocked progression towards widespread adoption, but which are rarely discussed in the literature. We argue that without substantial resources being focused on these issues, machine learning will continue to see limited application in clinical radiology.

REPRESENTATIVE DATA AND EVIDENCE

In order to make an informed decision on whether to invest in machine learning technology, representative clinical performance results are required. However, such data are often severely lacking. Firstly, algorithm development and testing is frequently carried out on limited datasets, which may not be representative of the clinic and may not be associated with adequate “gold-standard” diagnoses. Many researchers use publically available data sets, such as the Alzheimer’s Disease Neuroimaging Initiative (ADNI) databases (www.adni.loni.usc.edu). However, these data sets are often acquired under research rather than clinical protocols, using legacy equipment. Thus, there may be substantial uncertainties on the likely performance for clinical data. For example, in the case of ADNI consider the CAD Dementia challenge. Participants were invited to create a classification algorithm based on only very limited training data (n = 30). All but 2 of the 29 submitted classification algorithms used ADNI data to supplement the training data. However, in almost every case the estimated algorithm accuracy from training data was lower than that on previously unseen test data.

Furthermore, machine learning algorithms are rarely tested in the environment for which they were notionally designed. Standalone performance results are usually generated in a controlled setting with no human interaction. However, results derived from machine learning tools will at some point need to be interpreted by clinicians. This is particularly pertinent for assistive reporting software, which is designed to directly influence the radiologists’ final decision. Without evidence of impact on the whole
reporting workflow, benefits to the health service and patient care cannot be predicted.

**Regulations**

When software which is designed to have an impact on patient diagnosis or treatment is released on to the market, it is usually subject to some form of regulation. In Europe, the manufacturer must adhere to the Medical Device Directive. Whatever the classification under the regulations products need to be designed in such a way that patient safety is not compromised and that testing is carried out to ensure that the product performs as intended. Ensuring patient safety may be difficult for certain “black box” algorithms, such as neural networks, where outputs and risks may be challenging to predict.

Most regulatory regimes require clinical trials to confirm the performance of the final design. Ongoing surveillance is usually required in order to identify and fix any bugs associated with the software. Thus, meeting regulations usually requires significant financial resources, and diverse expertise. Costs are generally higher if the risk classification is high, which may be the case for machine learning algorithms designed to directly impact radiologists’ decisions.

Furthermore, the current model of device regulation in Europe (and in other jurisdictions) assumes that medical products are static entities, with any substantial changes to the product requiring reapproval. For machine learning algorithms that are designed to continually relearn and adapt their outputs in clinic, this approach to medical device regulation is impractical.

**Health economics**

When deciding on whether to invest in particular medical products many healthcare systems utilise economic analysis to inform their decision. In the UK, the National Institute for Health and Care Excellence (NICE, www.nice.org.uk) places strong emphasis on such data when generating guidance on medical technologies. This ensures that developed products have survived a cost–benefit analysis, providing evidence that can facilitate widespread adoption in the clinical community. However, even for the most simplistic economic analysis methods, such as cost–consequence analysis, evidence is required to quantify resource implications of the technology, as well as data on the likely clinical benefits. For many machine learning algorithms, this is difficult. For instance, gathering convincing data on the implications for patient pathways of a computer-aided detection algorithm, as compared to standard reporting methods, is likely to require extensive testing with radiologists under realist clinical scenarios. Once again, this is likely to be expensive, complex and time-consuming.

**Heterogeneity of the clinical environment**

Machine learning tools cannot be implemented in isolation. If machine learning is to be used routinely, software needs to be integrated within the hospital infrastructure such that it can be easily accessed and used by reporters, according to local preferences, and data can be transferred to and from the analysis package as required. However, there are substantial differences between hospitals in terms of information technology resources, associated restrictions and clinical protocols and workflows. The perils of ignoring local circumstances are reflected in the recently reported failure of IBM Watson for Oncology to achieve widespread clinical adoption, with the algorithms’ perceived in-built bias towards the American healthcare system cited as a major reason for lack of sustained uptake outside the United States. However, designing software that is adaptable to many different settings is difficult and, ultimately, it may not be possible to accommodate all the requirements of different hospital environments.

**Data ownership**

Machine learning research often relies on the use of retrospective patient data, acquired as part of standard care procedures. The steps necessary to achieve ethical approval in such circumstances are well established in Europe and the United States. However, if patient data are used to train an algorithm that is then sold commercially for profit, issues around data ownership and ethics can arise, particularly when data were originally acquired by a state-funded healthcare system. Furthermore, if the data are acquired in Europe and is not fully anonymised, the General Data Protection Regulations (GDPR) apply, which requires that the processing of personal data is in line with one of the specified lawful bases. If consent is chosen as the lawful basis, individual patients must opt-in to allow use of their data for machine learning development. This is another hurdle to development and obtaining and managing such consent adds extra costs to the development process.

**Support and promotion**

As highlighted by a recent Kings Fund report on adoption of innovation in the NHS, significant investment is usually needed to promote and support implementation of new technology. Simply generating evidence of impact is not enough to guarantee uptake in a healthcare system. If machine learning is to become a truly game-changing technology in radiology, support is likely to be needed from IT specialists, managers, radiographers as well as radiologists to ensure it is properly integrated in clinic. Not only does this require protected time (and therefore increased financial support) but the end users and patients have to be persuaded of its’ merits. Significant investment is therefore also required to promote the technology, to ensure that clinicians actively push the implementation. However, the perceived threat to radiologists’ role from machine learning, which is often inflated by articles in the popular press, is likely to make it harder to persuade the clinical community of the need for change.

**Reflection**

Machine learning promises much but given all of the above considerations, it is clear that the resources required to push machine learning technology into the clinic are substantial. Furthermore, it will take more than increased finances to enable machine learning to deliver on its promises.

In recent years, there has been some recognition that the challenges of implementation need to be addressed, rather than continually focusing on development of the algorithms themselves. For instance, the UK government is seeking to implement
recommendations from the life sciences industrial strategy, which references adoption of artificial intelligence and the need for funding to help move technology beyond the research arena. There is also recognition in the document that issues around data ownership, legislation and economic evaluation processes need attention if widespread implementation is to become a reality. In the United States, the FDA has recently taken a more active role in trying to streamline regulatory approval for software, as laid out in the Digital Health Innovation Action Plan.

Another positive sign is that the FDA recently approved the first medical device that can diagnose disease without input from a clinician (IDx-DR). However, despite these new developments and initiatives the translational burden placed on new machine learning technology remains relatively unchanged.

There are some applications where the benefits from machine learning are likely to be so large that there will be sufficient backing from a multitude of sources to overcome all the challenges described (assuming the technology is sufficiently mature). For example, cancer screening examinations of the breast and lung generate large volumes of imaging data that human reporters must examine. Development of a computer system that can screen such images automatically would save a significant amount of money and reduce the pressure on radiologists’ time, giving a strong incentive for adoption. The potential market for developers of such software would be large, encouraging commercial investment.

Furthermore, there are some machine learning applications associated with lower risk activities, such as automated segmentation of tumours, where the barriers to adoption (particularly in terms of regulation) are likely to be less substantial.

However, we argue that for the majority of radiological applications the balance between potential benefits and likely costs is currently weighted too heavily in favour of costs so that widespread clinical adoption is unlikely to be achieved. A fundamental change is required if this situation is to be improved.

Perhaps the biggest issue facing machine learning developers (particularly those in smaller companies) is a lack of access to realistic clinical data. However, widespread sharing of patient data requires investment in infrastructure and is associated with significant reputational risk for the health provider (as demonstrated by negative publicity around projects such as NHS England’s failed Care.data programme). Therefore, in accordance with the UK government’s recent statement on artificial intelligence, the authors advocate the establishment of data trusts to create data sharing systems, and to control data flows in a secure, ethical and transparent manner.

Without such actions, there is a danger that the obstacles to routine application of machine learning throughout radiology will be insurmountable.

**FUNDING**

Jonathan Taylor is funded by a Doctoral Research Fellowship from the National Institute for Health Research (and Health Education England), Funder-id http://dx.doi.org/10.13039/501100000272, Award-id: HCS DRF-2014-05-002.

**REFERENCES**

1. Liew C. The future of radiology augmented with artificial intelligence: A strategy for success. *Eur J Radiol* 2018; 102: 152–6. doi: https://doi.org/10.1016/j.ejrad.2018.03.019

2. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. 2017. Available from: https://arxiv.org/abs/1711.05225 [cited 2018 Aug 6].

3. Bron EE, Smits M, van der Flier WM, Vrenken H, Barkhof F, Scheltens P, et al. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDEntia challenge. *NeuroImage* 2015; 111: 562–79. doi: https://doi.org/10.1016/j.neuroimage.2015.01.048

4. Ross C, Sweeney L. IBM pitched Watson as a revolution in cancer care. It’s nowhere close. Stat [Internet]. 2017. Available from: https://www.statnews.com/2017/09/05/watson-ibm-cancer/ [cited 2018 Oct 3].

5. Kostkova P, Brewer H, de Lusignan S, Fottrell E, Gildacre B, Hart G, et al. Who owns the data? open data for healthcare. *Front Public Health* 2016; 4: 7. doi: https://doi.org/10.3389/fpubh.2016.00007

6. European Union. Regulation (EU) 2016/679 of the european parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* 2016; 1–88.

7. Collins B. Adoption and spread of innovation in the NHS [Internet]. 2018. Available from: https://www.kingsfund.org.uk/publications/innovation-nhs [cited 2018 Jan 22].

8. Bell J. Life Sciences Industrial Strategy - A report to the Government from the life sciences sector [Internet]. 2017. Available from: https://publications.parliament.uk/pa/id201719/ldselect/ldsctech/115/115.pdf [cited 2018 Jun 8].

9. U.S. Food and Drug Administration. Digital Health Innovation Action Plan [Internet]. U.S. Food and Drug Administration. 2017. Available from: https://www.fda.gov/downloads/medicaldevices/digitalhealth/ucm568735.pdf [cited 2018 Oct 3].

10. Department of Health, Department of Business, Energy and Industrial Strategy. Making a reality of the Accelerated Access Review: improving patient access to breakthrough treatments [Internet]. 2017. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664685/AAR_Response.pdf.