THE ROYAL SOCIETY

The Academy *of* Medical Sciences

AI in health and care: from bench to bedside

Note of discussions at a Royal Society and Academy of Medical Sciences workshop
29 March 2019

The Royal Society and Academy of Medical Sciences

The Royal Society is the UK's national academy of sciences. The Society's fundamental purpose, reflected in its founding Charters of the 1660s, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Academy of Medical Sciences is the independent body in the UK representing the diversity of medical science. Its mission is to promote medical science and its translation into benefits for society. The Academy's elected Fellows are the United Kingdom's leading medical scientists from hospitals, academia, industry and the public service.

In April 2017, the Royal Society published the results of a major policy study on machine learning. This report considered the potential of machine learning in the next 5 – 10 years, and the actions required to build an environment of careful stewardship that can help realise its potential. Its publication set the direction for a wider programme of Royal Society policy and public engagement on artificial intelligence (AI), which seeks to create the conditions in which the benefits of these technologies can be brought into being safely and rapidly[1].

As part of this programme, in March 2019 the Society and Academy of Medical Sciences convened a workshop for researchers and clinicians on the application of AI in health and care. This contributed to a programme of Royal Society activities on *AI for social good* and an Academy of Medical Sciences project on *AI and health*. The meeting also followed a roundtable the Academy held earlier in the year, jointly with the Medical Research Council (MRC) and the National Institute for Health Research (NIHR), which explored the challenges, opportunities and priorities for research and development of AI-based data analytics in the healthcare sector[2].

This note summarises discussions at the workshop. It is not intended as a verbatim record and its contents do not necessarily represent the views of all participants at the event, or Fellows of the Royal Society or Academy of Medical Sciences.

Understanding AI: current abilities and near-term opportunities and challenges

The term 'artificial intelligence' refers to a suite of technologies and methods that allow computer systems to carry out tasks that would typically require some form of intelligence in humans. Today's AI is dominated by data-enabled methods, such as machine learning. Machine learning allows computer systems to learn how to perform a task through repeated exposure to examples or data.

Many current approaches to AI have their roots in methods developed in the 1980s. Their analytical power has grown rapidly in recent years as a result of:

- Changes to the data environment: data generation and acquisition is becoming easier, following advances in digital measurement technologies, large-scale investments in health-relevant data infrastructures such as BioBanks, and the ability to combine data from multiple sources, such as genomic data, images, health records, and wearables.
- More powerful computing facilities: increased raw computing power can support more powerful algorithmic approaches.
- Policy mechanisms to support data access and use across different health data environments, including NHS digitalisation programmes.

1. Further information on this programme is available at: https://royalsociety.org/topics-policy/data-and-ai/artificial-intelligence/
2. Academy of Medical Sciences, Medical Research Council and National Institute for Health Research (2019). AI and health. https://acmedsci.ac.uk/AI-and-health

Today's AI is well-placed to add value to a range of healthcare domains, especially in areas where:

- A large quantity of real-world data is both available for analysis and necessary in order to make effective decisions;
- The learning environment is well-defined and stable, and programmes can define a clear objective for the AI system to pursue; and
- Where data has features that are hard to define, making hand-crafted models to analyse those features difficult to build.

There are already examples of the successful application of AI to healthcare challenges, notably in the use of image analysis to diagnose the presence of certain diseases, for example in breast cancer and diabetic retinopathy[3]. These successes have contributed to excitement across the public and private sectors about the transformative potential of AI in health. In the near-term, new AI-enabled systems or applications could include:

- New applications of image analysis to detect and diagnose disease;
- The application of AI to find efficiencies in the operational management of healthcare facilities, including scheduling appointments, checking systems, managing waiting times, and reducing the frequency of 'do not attend' events;
- The use of wearables and consumer devices to manage personal wellbeing, especially amongst an aging population; and
- The development of natural language processing techniques to analyse electronic health records, creating new insights into co-morbidities or better understandings of drug efficacy.

AI for health and care: snapshots of its potential

Contributing to clinical trials

To understand the efficacy of different treatment approaches, clinical trials can use data from participants self-reporting the impact of treatments on their quality of life. However, self-reporting has well-characterised limitations, for example, there is often lots of noise in such data, or it is inaccurately reported. If more accurate measures could be created, it might be possible to generate more robust findings in future.

In one example of a project seeking to achieve this, wearable devices are being trialled to capture lifestyle data about trial participants, including the time spent walking, driving, and sitting. In clinical trials assessing treatments for conditions such as Duchenne muscular dystrophy, such data can give important insights into the extent to which different approaches support patients to retain their mobility.

It is possible to collect data from sensor-based devices that monitor mobility. However, this tends to be 'messy', with noise from multiple movements or different inputs. In order to extract insights from such data, it is first necessary to identify the characteristics of the data that relate to the behaviours that are of interest – in this case, walking, driving, or sitting. AI-enabled analysis is potentially well-suited to this task: it can identify patterns in data from different wearables, and these patterns can inform analysis of the quality of life impacts of different treatments[4]. As a result, AI could help develop more useful, patient-centric assessments of treatments.

While AI may be able to assist with the development and conduct of clinical trials, it is unlikely that AI-enabled analysis will ever be able to replace clinical trials. Even with large volumes of data it will be impossible to define all the pre-treatment factors that determine patient outcomes to formally allow for these in any analysis. Indeed, there are high-levels of unexplained variation in disease outcomes even for homogenous groups of patients, and the impact of interventions may be dependent on both individual risk factors and the ways in which treatments are administered at a local level. Unless this unexplained variation could be reduced, it is unlikely that AI would be able to detect meaningful patterns or results. Randomised controlled trials are and will remain central to evaluating the causal effect of different interventions, and to the development of safe and effective treatments.

3. See, for example: Beck A *et al*. 2011 Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Sci. Transl. Med. 3, 108. (doi: 10.1126/scitranslmed.3002564); Gulshan V *et al*. 2016 Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 312, 2402–2410. (doi: 10.1001/jama.2016.17216)
4. For further information, see: https://www.ctti-clinicaltrials.org/files/usecase-duchenne.pdf and https://www.ctti-clinicaltrials.org/files/novelendpoints-recs.pdf

Improving diagnosis, treatment, and healthcare delivery

Sepsis

Sepsis is one of the leading causes of death across the world, and the main cause of mortality in hospitals. As disease progresses, it affects multiple organ systems, including cardiovascular, respiratory, neurological, renal, haematological, intestinal and hepatic systems. Therapy for septic shock relies on the early administration of antibiotics, alongside support for the function of the aforementioned systems. This can include respiratory interventions and oxygenation, monitoring of potassium levels in the kidneys, or cardiac interventions. Different treatment strategies may be necessary depending on disease progression, and treatment often needs to manage complex patterns of interactions between organs.

To help navigate this complexity, a US study has developed an 'AI clinician' – based on the analysis of 17,000 previous cases – to make recommendations about the optimal treatment strategies for sepsis patients in intensive care[5]. These recommendations were then tested on a further 79,000 case records that the system had not previously encountered. Researchers found that the system generally recommended lower doses of intravenous fluids and higher doses of vasopressors than patients received in reality, and there appeared to be evidence to suggest that these approaches were associated with lower levels of mortality.

While this study points to ways in which AI might help support clinical decision-making, it has not yet been tested in a clinical setting, and its results – based on US cases – might not be transferrable to other healthcare environments or localities. However, it does highlight how AI can be used to search for patterns in vast numbers of cases and generate hypotheses based upon any patterns found.

Breast cancer

Over 150 patients are diagnosed with breast cancer in the UK every day, and 500,000 people world-wide die of this disease each year. While screening is an important tool in tackling this disease, thousands of cases are not identified by mammograms, cancers can develop between screenings, and they can result in false alarms. AI offers the possibility of optimising breast cancer detection using digital X-ray technology: risk prediction algorithms are producing promising results in improving the accuracy of predicting the presence of breast cancer from mammograms.

For such screening to work well for all patients, the algorithms require training data from diverse populations. One approach to this – currently being pursued by a UK-based study – is to develop international research collaborations that access data from multiple localities when developing AI systems[6].

Pre-term birth

Pre-term birth is the delivery of babies before 37 weeks of gestation. It affects 1 in 10 births worldwide, and is the leading cause of death in children under 5[7]. Medical and lifestyle interventions can reduce the likelihood of pre-term birth, if risk indicators can be detected in time.

Electrohysterography (EHG) is a diagnostic technique that analyses electrical activity in the uterus, and can be used to understand whether 'true' labour is likely within seven days. New studies are seeking to understand whether this technique might be applied earlier in pregnancy, in order to predict whether a pre-term birth is likely.

By analysing large amounts of data from term and pre-term births, a machine learning system was able to detect and predict the likelihood of pre-term birth with greater sensitivity than existing techniques. Creating the carefully-curated datasets necessary to perform such analysis can be a challenge: for example, in one study an open dataset of 300 records used for analysis contained only 38 pre-term entries, and this disparity needs to be accounted for in curating the data set and preparing data for analysis[8].

5. Komorowski, M. *et al* (2019) The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care, *Nature Medicine*, 24, 1716 – 1720
6. For example: https://deepmind.com/blog/breast-cancer-screening-japan/
7. https://www.who.int/news-room/fact-sheets/detail/preterm-birth
8. Fergus, P. *et al* (2013) Prediction of preterm deliveries from EHG signals using machine learning. PLOS 1, 8 (10), https://doi.org/10.1371/journal.pone.0077154

Supporting social care

In many countries with aging populations, there is hope that AI-enabled robotics might offer support for patients and carers in allowing people to live healthier and independent lives for longer.

The development of assistive robots offers ways of supporting people to live independently in ways that are not currently possible. A range of assistive robotics already exist, or are in various stages of development, including, for example, emotional support robots to help patients deal with anxiety or physical assistant robots to help people with daily tasks.

If these devices are to be enthusiastically adopted, they will need to be able to communicate with people in ways that encourage positive interactions. One element of these positive interactions might be the ability to take into account cultural sensitivities or preferences of their users.

The CARESSES[9] project is seeking to design – and evaluate the impact of – assistive robots that are configured to interact with care home residents in a culturally-aware manner. Using AI, these robots learn the culture, customs, and etiquette of the person they are assisting, and use this learning to personalise interactions.

From bench to bedside: the application of AI in health and care

Data access and use

Current AI approaches perform well in areas where well-curated data is available for analysis. The NHS is the largest single health payer system in the world, and its data could be a valuable asset for the UK in seeking to develop healthcare applications of AI, if its data infrastructures can be developed and deployed strategically.

Over the last decade, there have been a number of programmes seeking to digitise NHS records and processes. NHS Digital's *Paper Free at the Point of Care* programme, for example, has encouraged a shift towards digitisation, with 500 million e-prescriptions now issued each year, and streamlined processes for clinical assessments. However, the landscape of NHS digitalisation and data management remains complex:

- The 'digital maturity' of systems across the country is variable, and further work is needed in many areas.
- There are many different types of health record, at varying stages of digitalisation.
- There are over 200 commissioning groups, across many different healthcare trusts, which might each have different approaches to data management.
- Links between datasets are inconsistent, and often missing between health and social care datasets.
- Data is often not held in an easily-analysable format – it might be split across worksheets or consist of different formats which are difficult to join together for analysis.
- In some cases, healthcare professionals might be reluctant to provide access to data, owing to concerns about data privacy and confidentiality.
- At a local level, GP data tends to be held by a small number of providers, which each have different policies towards data access and use. This can make access to primary care data very costly for researchers.

In addition, the data itself is often 'messy': it might have missing entries, have unclear provenance, or suffer from selection bias. While disciplines such as epidemiology and biostatistics have well-established ways of managing such data issues, these have not yet been translated across to AI research.

While many of these are ongoing challenges, there are a number of initiatives, led by the NHS and others, aimed at resolving these issues to produce a digital NHS where data are linked, interoperable and accessible for care and research.

9. http://caressesrobot.org/en/

There are also signals of a growing movement amongst patients to support researchers to access healthcare data, for example through the use of apps that provide data for use in research. While promising, these initiatives potentially introduce bias into the datasets collected in this way, as those patients engaging with the apps might not be representative of the wider population.

One response to these challenges might be the creation of civic data trusts. These would provide spaces for citizens to combine data and provide access to trusted partners, in order to facilitate research progress in key areas.

Supporting healthcare professionals to work alongside AI

One hope for the future development of AI is that these systems could create efficiencies or support processes that help address skills shortages in the delivery of health and care. The UK currently has 110,000 job vacancies in social care, for example, and AI-enabled systems might help ease some of the pressures in this system.

An important consideration in the design of AI systems is the way in which these applications might work alongside people. In cancer diagnosis, for example, there is a demand for skilled healthcare workers who are able to analyse scans to identify cancerous features. While this is an area of skills need, and AI systems might be able to address some of that need, it is not yet clear exactly what functions AI could fulfil, which functions patients would accept AI carrying out, and whether the use of AI could contribute to a broader de-skilling of people in the system. How these technologies fit into the wider healthcare system is therefore key.

Bringing people with domain expertise into the process of developing AI-enabled healthcare tools is also important to ensure the effectiveness of those tools. Those working in clinical settings often have contextual understandings that can explain features of the data or why different treatment approaches might work (or not) in different contexts, and such knowledge can improve the design of AI systems and ensure they are deployed effectively. In order to enable such collaborations, healthcare professionals may need additional skills in data literacy, to enable them to contribute to the development and calibration of these tools, and to challenge the recommendations or predictions that these tools produce.

Trustworthiness and public confidence

As public, policy, and research debates about the application of AI progress, a range of questions about trust follow: How can clinicians have confidence in the algorithms they interact with? How can research help create trustworthy AI systems? And how can researchers engage effectively with patients, creating a research environment in which patients feel they are benefiting from advances in the use of AI?

Developing robust and reproducible AI

Unlike traditional statistical approaches to data analysis – where the parameters of models are explicitly and well-characterised, and where levels of uncertainty can be analysed and assigned values – AI methods are primarily focussed on prediction. This creates particular challenges surrounding the explainability, reproducibility, and robustness of some AI systems.

As AI moves from being a research domain to one applied at scale, care is needed to ensure that users can be confident in the predictions coming from these systems. For example, clinicians will need to be confident that a system for operational management of a hospital designed in one area can be applied successfully in another.

While AI systems are able to produce accurate predictions, some methods – for example, those based on deep learning – are difficult for human users to explain. As systems based on 'deep learning' process data, the features of that data are used to strengthen or weaken connections between nodes across a neural network[10]. There can be complicated patterns of connections between these nodes, with potentially many thousands of layers and connections. Such systems can create highly accurate results, but it is difficult to explain why a result has been obtained.

While detailed technical explanations might not be helpful for all users, in the absence of a specific model – or the ability to falsify a hypothesis – it can be difficult for developers to reproduce results reliably. This leads to questions about the robustness of the analysis from some types of AI system.

10. A neural network is an approach to machine learning in which small computational units are connected in a way that is inspired by connections in the brain. These systems may consist of many layers of 'neurons': the base layer receives an input from an external source, then each layer beyond it detects patterns in activity from the neurons in the layer beneath, integrates these inputs, and then passes a signal to the next layer. In this way, signals can be passed through many layers, before reaching a top layer where a decision about the input is made. So, if the initial input is an image, the initial signals might come from the pixels of the image, and the top-level decision might be what object is in the image.

One effect of this is that these systems can be vulnerable to so-called adversarial attacks, in which a small, targeted change to the input data is deliberately deployed in ways that cause a large change to the prediction that the systems make. In one example of such an attack in a system designed to diagnose the presence of skin cancer, a small change to input data was shown to change both whether the system classified a mole as benign or malignant, and its level of confidence in that prediction[11].

Awareness of the importance of reproducible and robust research in AI is growing. Some guidelines already exist to support this[12] and regulatory bodies are taking action to adjust current governance mechanisms to take into account the questions raised by AI. For example, the US Food and Drug Administration has published guidelines on the use of devices in clinical trials. In some areas, such as direct-to-patient apps, additional steps may be necessary to ensure that patients can be confident in the use of these systems.

While careful study design and use of traditional statistics can help address these concerns, new tools or approaches, as well as cultural changes in the AI research community, may be necessary to develop more robust and reproducible approaches to AI. This might include codes of conduct or other methods for clinicians and AI researchers to work together in evaluating the effectiveness of AI-enabled tools.

Building a well-founded public dialogue

A side-effect of the rapid recent advances in AI, and its early successes in healthcare research, has been growing hype about its near-term capabilities, and their implications for society. This hype could lead to unrealistic expectations about the abilities of AI systems, the efforts required to develop AI that works in healthcare settings, and the timeline to deployment. It could also skew public debate, potentially putting public confidence in the technology and its applications at risk.

Amongst researchers, clinicians and policymakers there is an emerging consensus that public engagement and dialogue should be central to the design and deployment of AI systems for use in health and care. Recent reviews from a range of organisations have stressed the need for effective and meaningful public and patient engagement, and co-development of AI applications[13]. The challenge now is to create structures and processes that embed meaningful dialogue in the process of AI-enabled innovation.

In one example of efforts to create such structures, a 2019 Consensus Statement on Public Involvement and Engagement with Data-Intensive Health Research[14] set out guidance on the features of effective public dialogue. This noted the importance of ongoing engagement, with a clear purpose, that involves two-way communication, and is accessible to broad publics. Other initiatives have also noted the importance of well-informed and accessible information about AI and its applications for publics and communicators.

Building a well-founded public dialogue about AI technologies will be key to continued public confidence in the systems that deploy AI technologies, and to realising the benefits they promise in health and care. If the field is to make further progress in building this dialogue, it will be important for researchers working in AI and in medical sciences to have resources and support to co-design new studies with members of the public.

11. Finlayson, S. *et al* (2019) Adversarial attacks on medical machine learning, Science, 363, 6433, p1287
12. Examples of such initiatives include AI-TREE; TRIPOD; CONSORT
13. See, for example, recent reports by the Royal Society (2017) *Machine learning: the power and promise of computers that learn by example*, and *Portrayals and perceptions of AI and why they matter*, available at https://royalsociety.org/topics-policy/data-and-ai/artificial-intelligence/ and recent reports by the Academy of Medical Sciences (2018). *Our data-driven future in healthcare* https://acmedsci.ac.uk/file-download/74634438 and Academy of Medical Sciences, Medical Research Council and National Institute for Health Research (2019). *AI and health*. https://acmedsci.ac.uk/AI-and-health
14. https://ijpds.org/article/view/586