

## PERSPECTIVE



## AI Applied to Cancer

# Fair shares: building and benefiting from healthcare AI with mutually beneficial structures and development partnerships

Richard Sidebottom<sup>1,2</sup>✉, Iain Lyburn<sup>1,3,4</sup>, Michael Brady<sup>5</sup> and Sarah Vinnicombe<sup>1,6</sup>

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Artificial intelligence (AI) algorithms are used in an increasing range of aspects of our lives. In particular, medical applications of AI are being developed and deployed, including many in image analysis. Deep learning methods, which have recently proved successful in image classification, rely on large volumes of clinical data generated by healthcare institutions. Such data is collected from their served populations. In this opinion article, using digital mammographic screening as an example, we briefly consider the background to AI development and some issues around its deployment. We highlight the importance of high quality clinical data as fundamental to these technologies, and question how the ownership of resultant tools should be defined. Though many of the ethical issues concerning the development and use of medical AI technologies continue to be discussed, the value of the data on which they rely remains a subject that is seldom considered. This potentially controversial issue can and should be addressed in a way which is beneficial to all parties, particularly the population in general and the patients we serve.

*British Journal of Cancer* (2021) 125:1181–1184; <https://doi.org/10.1038/s41416-021-01454-2>

## INTRODUCTION

Artificial intelligence (AI) is generally reckoned to date from the 1950s, the term being coined originally by Professor John McCarthy. It developed strongly in the 1960s and 1970s, concentrating primarily on symbolic, as opposed to numerical, computation, with particular foci on natural language understanding and on reasoning. Until the mid 1980s, machine learning (ML) was a small sub-field of AI, but impetus was given to ML by the development of back-propagation methods and hidden layers between input data and output results during the training phase. ML methods have since developed strongly, to the point where techniques based on deep learning are now sufficiently advanced to show impressive results in retrospective studies<sup>1,2</sup> and are beginning to be used in clinical practice.<sup>3</sup>

Deep learning aims to uncover relationships within the data that are latent, and appears to be a more successful approach to medical image analysis than previous methods of computer analysis such as explicitly programmed feature extraction. Deep learning may be combined with other more conventional filtering and analysis methods within an AI system. Indeed, recent advances offer a realistic chance to improve many areas of medical care and population health. The arguments raised here apply to many medical areas; this article focuses on breast screening mammography as an example.

Deep learning methodologies require two fundamental ingredients to produce a clinically useful model: the “architecture” of

the machine learning system; and the clinical data used to train, validate, and then test the model. Developing these technologies requires extensive input from both. In ML parlance, the “architecture” refers to the technical details of the structure of the ML system: the number of stages, the number of data inputs at each stage, the spatial resolution of the data at each stage, and the “loss function” that drives the learning. There are numerous books, tutorials, and freeware available that explain the current crop of choices and facilitate experimentation. Several of the most commonly used and successfully applied deep learning architectures are publicly available,<sup>4,5</sup> and software to realise popular architectures such as ResNet and U-Net are available in software libraries such as Python. However, the customisation, tuning and set up involved in adaptation to any given application all require significant investment in time, expertise and computing capacity. Since this article is primarily about the clinical data used for training, validation, and testing, we largely omit further discussion of the topic here. It suffices for the purpose of this article to note that, in most cases, ML systems, particularly deep (i.e. many stages) convolution (i.e. oriented toward images as opposed to unstructured data) neural networks require large volumes of data to train, validate, and test them. From the published literature, most groups claiming successful results have used tens of thousands or hundreds of thousands of examples for training and development.<sup>1,2,6</sup> The provision of such large, carefully curated medical data sets is often challenging. Note that by way

<sup>1</sup>Department of Radiology, Gloucestershire Hospitals NHS Foundation Trust, Gloucestershire, UK. <sup>2</sup>Department of Radiology, The Royal Marsden Hospital NHS Foundation Trust, London, UK. <sup>3</sup>Cobalt Medical Charity, Cheltenham, UK. <sup>4</sup>Cranfield University, Cranfield, UK. <sup>5</sup>Department of Oncology, Medical Sciences Division, University of Oxford, Oxford, UK. <sup>6</sup>University of Dundee, Dundee, UK. ✉email: richard.sidebottom@nhs.net

Received: 24 November 2020 Revised: 21 April 2021 Accepted: 28 May 2021

Published online: 14 July 2021

of contrast, in non-medical domains, such “very large” data sets are increasingly available via the internet, Cloud, and social media (where images are posted).

Health data, for example from screening mammography, together with associated metadata, for example demographic information and identified pathology, is available throughout the world. However, such data is not often available in a form that is immediately suitable for training a deep learning model. Curation of representative datasets of sufficient size and quality to enable ML development is demanding and requires planning and resources. Moreover, there are potential conflicts between, on the one hand, medical advances based on innovative use of patient data, which was not originally gathered for that purpose, and, on the other, maintaining public trust in patient data protection. Evidently, we must ensure public support for such novel uses of their data. One approach is to establish partnerships between trusted data holders and AI algorithm developers, developing ethical guidance and public information campaigns. We should follow an approach which is transparent and mutually beneficial whenever feasible. There are projects underway which aim to promote this.<sup>7</sup>

#### DATA REQUIREMENTS, CURATION AND VOLUMES

As noted above, deep convolutional neural networks require large volumes of data to reach the high levels of performance that are required for use in clinical practice. While it is reasonable to suppose that the performance gain of using additional data reaches a plateau, the data requirements to achieve such a plateau may be very high given the wide variation in features contained within medical data. No upper limit has yet been demonstrated. This raises the exciting possibility that the performance of a system, perhaps as a combination of expert human and AI, could eventually substantially exceed that of human interpretation alone. It is therefore likely that the most valuable data will be at a population scale rather than from an individual hospital or small group of hospitals.

The quality of the data on which the training is performed is crucial. “Ground truth” is a term used for the values on which AI systems are typically trained (supervised learning). The underlying truth, for example the presence or absence of cancer, may be difficult to establish; but, as in almost all medical research, an approximation to the truth must be based on the information available. In the case of breast screening, a pathologically confirmed cancer may be considered “truth”; but it is more difficult to establish that a mammogram is “normal”. Long term follow-up data allows characterisation of an earlier study as normal if no cancer has developed in the interim period. Inaccurate or incomplete records may render ground truth data unreliable, making it unusable, or even detrimental, for use in deep learning.

It is essential that the datasets used for training, validation and testing of a ML system should be representative of the variation contained within the target population. Such variety includes: technical differences in mammography equipment; ethnic diversity within a population, and should capture or encompass to some extent the wide variety of normal and pathological features represented on mammograms. Such representative data sets will inevitably be very large and will offer the potential to develop training data which can be highly tuned.

Using such data sets, it may be possible to prioritise detection of the cancers of greatest biological significance so that a system offers increased sensitivity to them, whereas lesions which are of lesser significance may be de-prioritised. For example, small high grade invasive cancers are considered to be the most important cancers to detect at mammographic screening because they are most likely to kill if left untreated. It is feasible to train a model specifically to detect such lesions. It may be beneficial to have a

more sensitive but less specific operating point for such cancers as compared to more indolent appearing lesions. In the case of breast screening, such an approach would be likely to favourably influence the balance of benefits and harms by increasing the number of significant cancers that are detected and by reducing the number of unnecessary investigations and treatments for lesions which are unlikely to cause harm.

However, small high grade cancers are relatively rare in a heterogeneous screening dataset. They are technically difficult to identify at mammographic screening and, because they tend to have a faster sojourn time, many will not be captured by bi- or triannual screening. They are more often identified when they are larger, or when they present symptomatically as interval cancers. This means that examples derived from population screening of millions of women would be required to capture this detailed granularity required for optimal training. Such granularity would help to make further advances in personalised screening, where investigations are tailored to the individual according to their risk of developing a harmful malignancy.

#### ADVANTAGES TO THE NHS OF CO-ORDINATING THE USE OF DATA

The UK NHS breast screening programme (NHS BSP) comprises a large amount of imaging data of relatively consistent high quality across the diverse populations of the UK. This suggests that it could be a powerful source for training and testing AI models. Though the programme only screens just over 2 million women per year, compared to the approximately 40 million women screened annually in the USA, the routine and essentially uniform collection not only of images but also of associated metadata potentially offers an advantage to AI developers.

The NHS holds large quantities of similar scientifically and commercially valuable datasets collected during routine clinical care. The availability of long term follow-up data allows more reliable determination of ground truth. However, at present, these datasets are distributed around different NHS trusts, compiled using a variety of information systems, and, without dedicated curation or smart information systems, are not readily available for use. Happily, there is substantial expertise within the UK research community to define and identify appropriate data points to determine ground truth and to develop high quality datasets from this wide variety of NHS systems to the point where they could be suitable for use in AI applications.<sup>8</sup> The use of cloud infrastructure addresses many of the challenges around data size, processing and potentially for methods of implementation. However, it is important to note that the value of this data is not perpetual, not least as imaging and other technologies evolve, downgrading the importance of legacy data compared to data acquired using contemporary techniques.

At present, the regulatory and governance procedures which must be navigated by AI developers wishing to work with the NHS BSP are in many cases laborious and time-consuming. The decision makers who control access to the data find themselves in relatively uncharted waters, and tend, unsurprisingly, to tread cautiously. There is a need for guidance, simplification, and for the data to be unified in safe and secure ways so that the women served by the NHS BSP will be the ultimate beneficiaries of the huge benefits that robust AI algorithms could provide. By taking the initiative on these developments the NHS central commissioning bodies may be able to influence the direction of technological development. They could, for example, stipulate the desired behaviour of models trained on their data by specifying the balance in sensitivity and specificity in a variety of scenarios, such as in the case of suspected small high grade cancers discussed above.

Different tasks for AI tools will also require different behaviours. For example, an AI system could be tasked with triaging out

normal screening mammograms, thereby freeing up the relatively scarce, over-stretched breast radiologists for those tasks that demand their expertise. This would require a very different operating point, with a high negative predictive value, compared to that where it is acting as a second reader. Dialogue between technology developers and service providers to better understand each others capabilities and needs, supported by appropriate data use, should enable improved tools to be developed.

Independent test sets are used to demonstrate the performance of an AI system for a defined task. Test sets should be highly representative of the target population and must not have been seen or controlled by technology developers. Such test sets, of sufficient size to offer reliable results, are urgently required for assessment of these technologies in retrospective studies. Subsequently, prospective clinical studies should be performed prior to widespread adoption. The NHS needs to avail itself of these exciting possibilities and important responsibilities. The recently launched NHSx project “National AI Medical Imaging Platform”<sup>9</sup> will hopefully be able to address exactly these issues. What the NHS does could positively influence the impact of these technologies in the UK and on other healthcare systems around the world.

### OWNERSHIP AND CONTROL OF TRAINED AI MODELS

Currently, in most instances, the default model for development and use of this type of clinical tool is for a technology developer to gain rights to use population/patient derived data and retaining all rights and control over the resultant model. However, any such model has had vital input not only from the engineers and scientists who have built them, but also from the clinical data on which they were trained—the individual patient’s information and the clinical team’s input in terms of image interpretation and subsequent management.

The ultimate impact that AI technologies will have on society will be profound. Optimal outcomes will only be achieved through mass participation. This will only happen if there is a high level of trust and a belief that this is a fair use of population data. A recent UK government survey<sup>10</sup> reaffirms that the NHS enjoys a high level of trust amongst the public for data handling, exceeding that even of banks and friends. This trust is valuable and must be cherished and protected. Participants in screening programmes implicitly consent to the use of data for research, but may not be explicitly aware that data may be used for developing commercial products.

Recently the UK government has published a policy paper “Preparing for the National Data Strategy”.<sup>11</sup> This has highlighted some of the issues covered above including data availability and the opportunities that this data offers. At the moment, no mechanism has been proposed to recognise the inherent value of the contribution of health data derived from the population, and this issue is seldom considered.<sup>12,13</sup>

Provided that absolute safeguards are in place as regards data protection, and so long as the main beneficiaries are their fellow citizens, the overwhelming majority of the population is likely to support the use of their data to improve the healthcare system.

The same government survey<sup>10</sup> showed that 79% of respondents agreed they would share data about themselves to develop new medicines or treatments. This is far higher than for any other purpose included in the survey. Widespread public support is less likely if there is the perception, real or otherwise, that the data is being harvested for the advancement of profit for commercial providers or used for purposes that are unexpected.

Perhaps an equitable way for society to maximally benefit from AI technologies is to have a stake in the trained models where population data has been used, as has been done in other parts of the European Union, notably the Netherlands. In particular, the population that contributed data must have access to any

improvements that this data enables through whatever system is envisaged.

To maximise the success and potential of the opportunity to enhance medical care and improve clinical outcomes, balanced even handed approaches in true partnership are likely to get the greatest support from the general population, health care systems and AI development teams.

### CONCLUSION

Large-scale well curated clinical datasets, such as could be built with mammographic breast screening data, will be essential to realise the benefits that AI techniques can bring to healthcare. If data remains isolated in hospital systems and is not used for the development of AI systems, everyone will miss out on the potential benefits. To maximise the potential of our data we should invest in building usable national datasets, allowing us scope to direct and lead these developments. However, there is a potential tension between maximising the scientific power of this data and recognising and protecting the contribution of individuals and society, along with the commercial value that their contribution represents. New approaches to addressing this balance are urgently required. If we can achieve this, we all benefit by allowing healthcare AI to flourish.

### REFERENCES

- 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.* **111**, 916–922 (2019).
- McKinney, S., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafiyan, H. et al. International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89–94 (2020).
- FDA. ScreenPoint medical FDA clearance. [https://www.accessdata.fda.gov/cdrh\\_docs/pdf19/K193229.pdf](https://www.accessdata.fda.gov/cdrh_docs/pdf19/K193229.pdf) (2020).
- He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. <https://doi.org/10.1109/CVPR.2016.90> (2016).
- Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI’17). 4278–4284 (AAAI Press, 2017).
- Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M. et al. Deep neural networks improve radiologists’ performance in breast cancer screening. *IEEE Trans. Med. Imaging* **39**, 1184–1194 (2020).
- NCIMI. About NCIMI. <https://www.medsci.ox.ac.uk/research/networks/national-consortium-of-intelligent-medical-imaging/about-ncimi> (2020).
- Halling-Brown, M. D., Warren, L. M., Ward, D., Lewis, E., Mackenzie, A., Wallis, M. G. et al. OPTIMAM Mammography Image Database: A Large-Scale Resource of Mammography Images and Clinical Data. *Radiol Artif Intell.* **3**, e200103. <https://doi.org/10.1148/ryai.20200103> (2020).
- NHSx. AI in imaging. Our work with databases. <https://www.nhs.uk/ai-lab/ai-lab-programmes/ai-in-imaging/> (2021)
- Department for Digital, Culture, Media & Sport. Opinions and lifestyle survey data module. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/962551/February\\_2020\\_Opinions\\_and\\_lifestyle\\_survey\\_data\\_module\\_V2.xlsx](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/962551/February_2020_Opinions_and_lifestyle_survey_data_module_V2.xlsx) (2020).
- Department for Digital, Culture, Media & Sport. Policy paper: National Data Strategy. <https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy#data-1-4> (2020).
- Vollmer, S., Mateen, B., Bohner, G., Király, F., Ghani, R., Jonsson, P. et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics and effectiveness. *BMJ* **368**, l6927 (2020).
- Gilbert, F., Smye, S. & Schönlieb, C. Artificial intelligence in clinical imaging: a health system approach. *Clin. Radiol.* **75**, 3–6 (2020).

### ACKNOWLEDGEMENTS

We thank Sarah Haden and Angela Haden for help with drafting and editing.

**AUTHOR CONTRIBUTIONS**

Conceptualised the review: R.S., I.L., S.V. Contributed to writing early drafts of the paper: R.S., I.L., M.B., S.V. Revised the final version of the paper: R.S., I.L., M.B., S.V.

**FUNDING**

No funding was provided for the preparation of this article.

**ETHICS APPROVAL AND CONSENT TO PARTICIPATE**

Not applicable.

**DATA AVAILABILITY**

Not applicable.

**COMPETING INTERESTS**

R.S.: paid consultant to Google health, previously DeepMind health; paid consultant to Micrima Ltd.; and advisory board for the Cancer Research UK OPTIMAM project

(unpaid). I.L.: Medical Director Cobalt medical charity; paid consultant to Micrima Ltd.; and advisory board for the Cancer Research UK OPTIMAM project (unpaid). M.B.: Founder, Chairman, Shareholder of Perspectum Ltd. (quantitative MRI for NASH, liver cancer, diabetes, COVID); Founder, Chairman, Shareholder of ScreenPoint by (decision support for mammography, tomosynthesis); Founder, Shareholder of Mirada Medical Ltd. (radiological image fusion, decision support for radiation therapy); Founder, Shareholder of Volpara Health (breast imaging services, density, risk); Chairman, Shareholder of Optellum Ltd. (Lung nodule analysis); Interim President, Mohammad Bin Zayed University of Artificial Intelligence, Abu Dhabi; and Emeritus Professor of Oncological Imaging, Oncology, University of Oxford. S.V.: President of the British Society of Breast Radiologists (no competing financial interests).

**ADDITIONAL INFORMATION**

**Correspondence** and requests for materials should be addressed to R.S.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.