AI in radiology — from algorithm development to implementation in clinical routine

Over the last few years, there has been a continuing, ever-increasing interest and debate over the role that Artificial Intelligence (AI) – based technologies can, and should, play in radiology. Now the debate is over — AI in radiology is here to stay, although in exactly what forms and applications remains to be defined. Many technological, validation and regulatory issues have still to be overcome.
To put the practical implications of AI in radiology into context, we spoke to Prof Declan O'Regan, Consultant Radiologist and Director of Imaging Research at Imperial College, London, UK.
Prof O'Regan combines clinical radiological work with research into the development of AI-algorithms, including industry-funded projects supported by Bayer.

Prof. Declan O'Regan is a consultant radiologist and Director of Imaging Research at Imperial College, London, UK.
email
declan.oregan@lms.mrc.ac.uk

Q Before we get into the broader subject of AI in radiology, can you please give us a brief summary of your activities at Imperial College in London

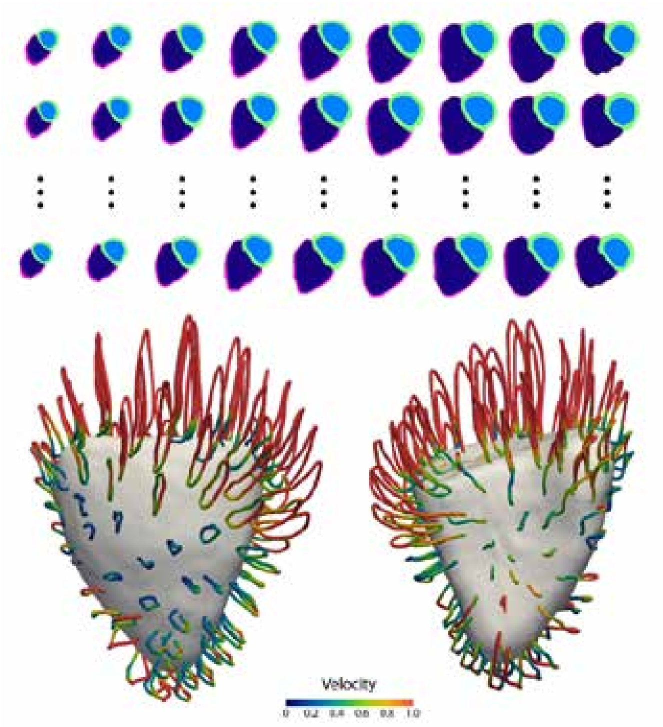
As your introduction suggests, my interest indeed straddles both the use of AI in clinical departments but also the development of new algorithms.

Let's start with our research activities. These are funded by the UK Medical Research Council (MRC), the British Heart Foundation (BHF), and collaborations on research and development with Bayer.
Our unit is physically located in the MRC London Institute of Medical Sciences (MRC-LMS) sited at the Hammersmith teaching hospital. The hospital is itself a part of the Imperial College Healthcare National Health Service (NHS) Trust, which provides acute and specialist healthcare in north west London for around a million and a half people.
The MRC LMS is a biomedical research institute where scientists and clinicians collaborate to advance the understanding of biology and its application to medicine, bridging basic science and clinical research.

In addition to the Hammersmith hospital where MRC-LMS is based, the NHS Imperial Trust also covers two other large teaching hospitals in London, namely St. Mary's and Charing Cross. In total the number of radiology examinations carried out throughout the Trust is huge: approximately 90 000 CTs per year, and 70 000 MRIs per year.
For our research in the MRC-LMS, we have two MRI systems, one 1.5T and the other 3.0 T which we use for carrying out innovative imaging research, all in the cardiovascular domain. This work includes exercise cardiac MRI, where we exercise people in the magnet, as well as advanced 3D cardiac imaging, 4D flow mapping, and spectroscopy analyses.
We also develop and test AI algorithms, which aim not to replicate tasks that radiologists perform, but to add value by increasing the amount of information that can be gained from medical imaging, particularly with the aim of improving prognoses. .

"... We also develop and test AI algorithms, which aim not to replicate tasks that radiologists perform, but to add value by increasing the amount of information that can be gained from medical imaging, ..."

New tools that help to better define diseases and risk groups at an early stage might help in offering better treatments. For this, we think it is important to understand the significance of genetic differences and how the heart develops so as to be able to identify the patient's risk of developing heart disease in the future.
As part of our research interests



Machine learning cardiac motion analysis of patients diagnosed with right heart failure is used for automated survival prediction.
Image reproduced from Bello *et al.* Deep learning cardiac motion analysis for human survival prediction. Nat Mach Intell. 2019; 1: 95-104.
https://doi.org/10.1038/s42256-019-0019-2.
The work described in the paper demonstrates how a complex computer vision task using high-dimensional medical image data can efficiently predict human survival.

we use data from other patient populations, such as via access to the UK Biobank, the large long-term biobank study investigating the respective contributions of genetic predisposition and environmental exposure to the development of disease. We also use other data held by the NHS, in essence involving the integration of large data sets. We do this in collaboration with several research groups throughout the world.

Our unit's mission statement summarizes nicely what we do: our overall focus is the exploration of mechanisms underlying heart function using machine learning (ML) to analyze cardiac motion for predicting patient outcomes, discovering potential therapeutic targets and identifying genetic risk factors.

Q *Has the work you're doing with Machine Learning already resulted in the production of algorithms that are now being used in clinical routine?*

We're not quite at this stage yet but we have developed an algorithm called 4Dsurvival which we are hoping to beta test here in London in the next twelve months. Basically the algorithm operates by analyzing the motion of the heart so that the clinician can then be informed of the level of risk that the patient has of developing heart failure in the future. The output of the algorithm is thus actionable information. If a patient is identified as being in a particular risk group, the treatment could be changed appropriately, according to guideline derived choice of treatment paths.

As I said, we'll soon be starting beta testing of the algorithm, although the ultimate proof of its efficacity will be through randomized controlled trials (RCTs). Of course, RCTs take time and are expensive but for clinical decisions that are based on AI algorithms, RCTs are the method that provides the strongest evidence as to how well they work in practice .

More generally, I think that, ML is increasingly maturing beyond the development of new algorithms and their validation but also to the carrying out of well conducted, well-controlled RCTs. We know real reproducibility can be an issue in ML. Frequently, algorithms are only validated in a relatively small number of cases with the result that the performance of the algorithm can be disappointing. This is where the clinical trials come in. Of course you can always have trials which are well designed and well conducted but neverteless the algorithm still disappoints when it comes to its use with real-world data of clinical routine.

For this reason it's really important to have continuous feedback on the performance of the algorithm in the field, if only to understand how algorithms may fail. In this analysis we shouldn't forget that not all algorithm errors are equal — some errors are more important than others.

It also shouldn't be forgotten that AI has no "common sense". Thus, situations can arise with AI that would never occur with a human observer because of a common sense error that is easy for the human to spot. As an example, a recent paper evaluated an algorithm designed to interpret plain film images for the detection of fractures. The algorithm was very accurate and worked very well — unless the patient happened to have a plaster cast, when the AI failed to identify any fractures. This was simply because such cases hadn't been present in the data set used to train the algorithm training. A human observer would have no difficulty in sorting such an issue out.

Thus, the evaluation of algorithms is more than just a question of measuring the headline sensitivities and specificities — it is also important to understand the interactions between humans and the algorithms. For example how much weight does the clinician really put on the outcome of an algorithm and how does it really affect the patient's clinical outcome? All these human factors are under-explored at the moment.

Q *If your algorithm was trained only on data from your unit, is there a danger that it may not be applicable more broadly?*

Imperial is one of several partner trusts (there are ten in total) in the London AI center, a part of whose mission is to develop federated learning facilities which allow algorithms to be trained on the more diverse data from within several trusts. In our implementation of federated learning, the data actually all stay within the appropriate trust where the data were originally generated, so maximizing data security and patient confidentiality aspects. Instead of the data leaving the Trust, the algorithm goes to the data. There are also an increasing number of trusted research environments which allow the iterative development of algorithms, where they can be fine tuned and their applicability to broader populations can be checked.

Q *You mentioned human factors earlier in our conversation. Broadening out to the reaction of radiologists in general, how are radiologists reacting to the onset of AI? Do they perceive it as an existential threat or a useful, even welcome help?*

There was a concern some years ago about AI adversely affecting radiologists' career prospects but no radiologists have ever had their roles taken over by a machine. In my opinion such a scenario is highly unlikely any time soon, beyond very specific use cases. You shouldn't forget that the negative comments made several years ago about the future of the profession as a result of AI came not from radiologists but from some computer scientists.
None of the fears expressed then have been borne out. On the contrary, the algorithms that are reaching maturity at the moment are appreciated by radiologists as decision support, improving the sensitivity and specificity in radiological exams as well as providing added value in areas such as work-list prioritization, etc.
When considering the impact of innovative technology, we shouldn't forget that radiology has always been a technology-driven specialty and has actually been at the forefront of adopting new technology and Machine Learning. The innovations can be potentially transformative, but I can't repeat often enough that it is essential to have strong, robust evidence that the tools perform as expected in practice, There is a lot more investigation still to be done to find out how these tools work in practice, and how radiologists interact with them.
In fact, in contrast to the gloomy predictions some years ago about the future impact of AI on radiology, I think that new technology will actually increase the role of imaging and the overall demand for images. An example of this is in the increasing use of the determination of Fractional Flow Reserve from CT images of the heart, CT_{FFR}. The arrival of this technology could even increase demand for CT examinations of the coronary arteries, which is quite understandable since CT_{FFR} gives more accurate information and may avoid unnecessary invasive coronary angiography. The fact that CT_{FFR} is recommended by the UK National Institute for Clinical Excellence (NICE) has also helped to support its implementation. Thus, if new technologies can enable radiologists to provide a more efficient service and to contribute more added value, I can see the role of radiology actually becoming ever-more pivotal in the overall patient's journey.

Q *If radiology has so much promise, how do you explain the current chronic shortage of radiologists or, as has been reported in some European countries a decline in the number of newly-graduated students of medicine choosing radiology as a specialty? Is AI responsible?*

In fact, in the U.K. at least I don't see a shortage of strong candidates applying to specialize in radiology. We still have excellent graduates coming through to apply to do radiology, particularly those who are keen on an academic career, These applicants are stronger than ever and radiology is still highly over-subscribed.
Where there clearly is a chronic shortage in the UK certainly but I believe also in other countries, is of qualified consultant radiologists. This has been confirmed by work force surveys, e.g. those carried out by the UK's Royal College of Radiologists (RCR)
There is no single strategy that can be used to address this problem, but there's no doubt that we as radiologists have to look towards innovative ways of working, for example by looking at extended roles for radiographers, for instance. I think ML could also have a role in improving efficiency. That is certainly the focus of some of the vendors in the field — not just for image recognition, but in areas such as smart use of resources, prioritization of work flow, such as by moving urgent patients up the work list.

Q *Which do you think are the applications most likely to benefit from AI?*

The most promising immediate applications of AI are likely to be in screening programs such as those in mammography or in chest. In such screening environments AI could flag normal cases, leaving the radiologist more time to deal with positive or more difficult cases. But in any case, the radiologist will still have to review all the cases. Urgent ones could be prioritized, but the radiologist will still have to issue a report.
However, as I suggested earlier, there are many promising applications of AI in radiology other than in terms of lesion detection. The danger is that we get over-focussed on lesion detection, when the potential of AI in radiology is in fact much broader.
AI applications which have nothing to do with image interpretation, such as algorithms that interact with other health informatics systems, with Electronic Health Records (EHRs,) and actively prioritize workload planning, could have the biggest effect of all.

Whatever the precise application, the future of AI in radiology looks not just promising, but more and more necessary.

"... The danger is that we get over-focussed on lesion detection, when the potential of AI in radiology is in fact much broader..."