

THIS WEEK

EDITORIALS

DATA Funders must explain benefits of management plans **p.286**

WORLD VIEW Use ocean floats to track regional carbon emissions **p.287**



GRIPPING Meet the robotic arm with a soft handshake **p.289**

AI diagnostics need attention

Computer algorithms to detect disease show great promise, but they must be developed and applied with care.

One of the biggest — and most lucrative — applications of artificial intelligence (AI) is in health care. And the capacity of AI to diagnose or predict disease risk is developing rapidly. In recent weeks, researchers have unveiled AI models that scan retinal images to predict eye- and cardiovascular-disease risk, and that analyse mammograms to detect breast cancer. Some AI tools have already found their way into clinical practice.

AI diagnostics have the potential to improve the delivery and effectiveness of health care. Many are a triumph for science, representing years of improvements in computing power and the neural networks that underlie deep learning. In this form of AI, computers process hundreds of thousands of labelled disease images, until they can classify the images unaided. In reports, researchers conclude that an algorithm is successful if it can identify a particular condition from such images as effectively as can pathologists and radiologists.

But that alone does not mean the AI diagnostic is ready for the clinic. Many reports are best viewed as analogous to studies showing that a drug kills a pathogen in a Petri dish. Such studies are exciting, but scientific process demands that the methods and materials be described in detail, and that the study is replicated and the drug tested in a progression of studies culminating in large clinical trials. This does not seem to be happening enough in AI diagnostics. Many in the field complain that too many developers are not taking the studies far enough. They are not applying the evidence-based approaches that are established in mature fields, such as drug development.

Many reports of new AI diagnostic tools, for example, go no further than preprints or claims on websites. They haven't undergone peer review, and might never do so. That would verify key details: the underlying algorithm code, and analyses of, for example, the images on which the model is trained, the physicians with which it is compared, the features the neural network used to make decisions, and caveats.

These details matter. For instance, one investigation published last year found that an AI model detected breast cancer in whole slide images better than did 11 pathologists who were allowed assessment times of about one minute per image. However, a pathologist given unlimited time performed as well as AI, and found difficult-to-detect cases more often than the computers (B. E. Bejnordi *et al.* *J. Am. Med. Assoc.* **318**, 2199–2210; 2017).

Some issues might not appear until the tool is applied. For example, a diagnostic algorithm might incorrectly associate images produced using a particular device with a disease — but only because, during the training process, the clinic using that device saw more people with the disease than did another clinic using a different device.

These problems can be overcome. One way is for doctors who deploy AI diagnostic tools in the clinic to track results and report them, so that retrospective studies expose any deficiencies. Better yet, such tools should be developed rigorously — trained on extensive data and validated in controlled studies that undergo peer review. This is slow

and difficult, in part because privacy concerns can make it hard for researchers to access the massive amounts of medical data needed. A News story on page 293 discusses one possible answer: researchers are building blockchain-based systems to encourage patients to securely share information. At present, human oversight will probably prevent weaknesses in AI diagnosis from being a matter of life or death. That is why regulatory bodies, such as the US Food and Drug Administration, allow doctors to pilot technologies classified as low risk.

“Many in the field complain that too many developers are not taking the studies far enough.”

But lack of rigour does carry immediate risks: the hype–fail cycle could discourage others from investing in similar techniques that might be better. Sometimes, in a competitive field such as AI, a well-publicized set of results can be enough to stop rivals from entering the same field.

Slow and careful research is a better approach. Backed by reliable data and robust methods, it may take longer, and will not churn out as many crowd-pleasing announcements. But it could prevent deaths and change lives. ■

Russian research

The sleeping bear of Russian science could finally wake — and China can show it how.

Vladimir Putin will hardly be remembered as a patron of science. Not for Putin the scientific philosophy of dialectical materialism that helped to drive research in the former Soviet Union and that remains influential among many of his contemporaries. His long rule over Russia, as both president and prime minister, shows that he is more inclined to line up with the nation's Orthodox Church. His 2016 choice of an ultra-conservative religious historian as science and education minister was no accident.

But Putin, who is expected to win another six years in power in the Russian presidential elections on 18 March, did not get where he is today without being able to play both sides. He acknowledges — and has often said — that Russia's poor research and development capacity is an obstacle to economic growth and prosperity. His clique of political cronies includes scientists and research administrators. And their lobbying has not been in vain. Russian science spending has palpably (if by no means fully) recovered in recent years from near-collapse in the 1990s.

Outsiders recognize this: international sanctions in response to Russia's occupation of the Crimea have spared East–West research collaboration. And Russia's demanding education system continues

CORRECTION

The Editorial 'AI diagnostics need attention' (*Nature* **555**, 285–286; 2018) gave an inaccurate description of the methods in a 2017 study. The model detected breast cancer in whole slide images, not mammograms.