

ILLUSTRATION BY SAM FALCONER

Another set of eyes for cancer diagnostics

Artificial intelligence's ability to detect subtle patterns could help physicians to identify cancer types and refine risk prediction. **By Neil Savage**

When a young girl came to New York University (NYU) Langone Health for a routine follow-up, tests seemed to show that the medulloblastoma for which she had been treated a few years earlier had returned. The girl's recurrent cancer was found in the same part of brain as before, and the biopsy seemed to confirm medulloblastoma.

With this diagnosis, the girl would begin a specific course of radiotherapy and chemotherapy. But just as neuropathologist Matija Snuderl was about to sign off on the diagnosis and set her on that treatment path, he hesitated. The biopsy was slightly unusual, he thought, and he remembered a previous case in which what was thought to be medulloblastoma turned out to be something else. So, to help him make up his

mind, Snuderl turned to a computer.

He arranged for the girl to have a full-genome methylation analysis, which checks for small hydrocarbon molecules attached to DNA. The addition of such methyl groups is one of the mechanisms behind epigenetics – when the activity of genes is altered without any mutation to the underlying genetic code – and different types of cancer show different patterns of methylation. Snuderl fed the results to an artificial-intelligence (AI) system developed by a consortium including researchers at the German Cancer Research Center in Heidelberg, and let the computer classify the tumour.

“The tumour came back as a glioblastoma, which is a completely different type,” Snuderl says. The new tumour seemed to be the result of radiation used to destroy the first cancer, and called for a different drug and radiation treatment plan. Treatment for the wrong cancer could have ill effects without actually destroying the cancer. “If I had finalized the case just on pathology, I would have been

terribly wrong,” Snuderl says.

The system Snuderl used is an early example of AI as a tool to diagnose cancer. NYU Langone’s Perlmutter Cancer Center received state approval to use its AI classifier as a diagnostic test in October 2019, and researchers around the world are developing similar systems to help pathologists diagnose cancer more accurately. The goal is to use AI’s ability to recognize patterns that are too subtle for the human eye to detect to guide physicians towards better-targeted therapies and to improve outcomes for patients. Some scientists are even applying AI to screening tests in the hope of identifying people with an increased cancer risk or catching the disease sooner.

The methylation method

The methylation-based classifier, developed by a consortium of dozens of researchers, was originally trained to sort medulloblastomas into subtypes. The German-led team eventually expanded the effort to cover all of the 100 or so known cancers of the central nervous system. When the initial results were published¹ in March 2018, the researchers made the classifier available online. Other researchers can upload methylation profiles and, in a few minutes, learn which subtype the cancer fits into. They also receive a confidence score that says how likely the result is to be correct. About 1,000 such profiles are uploaded each month, says Andreas von Deimling, a neuropathologist at the German Cancer Research Centre who was one of the project’s leaders.

Although Langone’s use of the test has been approved by New York state, the website notes that the classifier is still a research tool that has not been clinically validated. The classifier was originally trained using around 2,800 tumour samples, but since the website has been operating, that number has grown to around 60,000. “This is much more than a single pathologist sees in an entire lifetime,” von Deimling says. “By the sheer number of tumours we can now examine with this system, we find novel entities no pathologist has ever been able to define previously.” The system compares data to its reference list of tumours and places the profile into a group, but if it doesn’t quite match, the cancer gets a low confidence score. Pathologists examine the low-scoring samples, and if there are at least seven with the same methylation profile, they assign them to a new group and retrain the classifier. The classifier now recognizes about 150 different cancer entities.

The computer’s ability to spot those cancer types could cut hospitals’ error rates. In the initial study, the algorithm found that 12% of brain tumours had been misdiagnosed by pathologists. Snuderl says that NYU has similar error

rates of 12–14% among its patients. “That’s not an insignificant number of people that could benefit simply from having the right diagnosis,” Snuderl says.

Methylation profiling is expensive – typically, only large cancer research centres can afford it. So the scientists hope to find simpler biomarkers to identify the subtypes. If, for instance, they can discover differences that are visible by looking at stained tissue under a microscope, they can make the same level of diagnostic sorting available to the many hospitals that don’t have the resources for methylation profiling. “You can develop these markers only if you have the grouping correct in the first place,” von Deimling says.

Getting it right

Correctly diagnosing cancers in other parts of the body can also be difficult. Working out if a person has prostate cancer and whether that cancer is aggressive enough to need treatment or merely needs to be watched can be tricky (see page S2).

Most prostate cancers are diagnosed by taking biopsies from a standard set of locations on the prostate, but this can mean the actual cancer is missed. A newer approach uses multiparametric magnetic resonance imaging (MRI), in which different types of MRI scan are combined. But highly trained radiologists don’t always agree on what they’re seeing in the images, and those with less experience do even less well at identification. “To reach a certain level of expertise in radiology, particularly in this prostate-cancer MRI diagnosis, requires a lot of training,” says Kyung Hyun Sung, a radiologist at the University of California, Los Angeles. As a major prostate-cancer treatment centre, the university has a programme to train radiologists to read such images and boasts specialists with ten or more years of experience. But that is not the norm. “Community hospitals don’t have that training period or expertise in their ranks,” says Sung.

With those hospitals in mind, Sung is building an AI-based system called FocalNet to help physicians to better classify prostate cancer. To train the programme, Sung and his colleagues collected around 400 pre-operative MRI scans of people who were going to have surgery to remove their prostate. The researchers fed FocalNet a subset of the scans, along with the tumour’s Gleason score – a rating of malignancy, defined by pathologists who analysed the tissue after the prostate was removed. The system then looked for and learnt to spot patterns in the MRI scans that matched the pathology-based score.

The researchers then challenged FocalNet to provide a Gleason score for a new set of scans.

The computer found 79.2% of the clinically significant cancer lesions, as determined by pathology. A group of radiologists, each with at least 10 years of experience of reading more than 1,000 images annually, managed 80.7% – a difference deemed statistically insignificant².

“The machine will be consistent. It’s not going to have inter-reader variability.”

Currently, the value of a Gleason score derived from an MRI is limited because it is dependent on the skill of the radiologist interpreting the image. But that, says Sung, is when machine learning can help. “The machine will be consistent. It’s not going to have inter-reader variability.” With the help of a system like FocalNet, multiparametric MRI could be used even without experienced radiologists, leading to clearer diagnoses that can guide people to the right treatment.

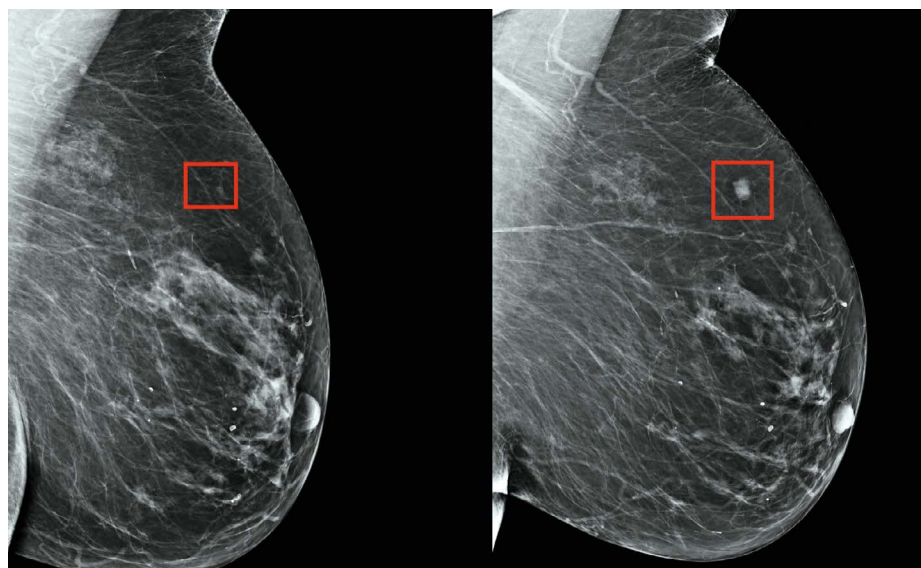
Screening for survival

Although getting the diagnosis right is important, catching cancer early can also lead to higher survival rates. Many women in the United States have annual mammograms starting in their forties or fifties. That produces a lot of imaging data. Regina Barzilay, a computer scientist at the Massachusetts Institute of Technology (MIT) in Cambridge, wanted to see if a machine could use those data to draw a more accurate picture of a person’s risk of developing breast cancer.

Barzilay collected almost 89,000 mammograms from nearly 40,000 women who had been screened over a 4-year period, and checked the images against a national tumour registry to determine which women were eventually diagnosed with breast cancer³. She then trained a machine-learning algorithm with a subset of those images and outcomes, before testing the system to see how well it predicted cancer risk. The computer put 31% of the women who eventually developed breast cancer into the highest risk group. But the standard Tyrer–Cuzick model that physicians use to estimate risk – based on factors such as age, family history of cancer, and age at first menstrual period and at menopause – placed only 18% in that group, even when physicians added measurements of breast density from mammograms to the model.

The researchers are continuing to improve the model, says Adam Yala, a PhD student at MIT who works with Barzilay on the project. The researchers hope that their work can lead

outlook



An AI system identified a woman's potential breast tumour four years (L) before it developed (R).

to more personalized breast-cancer screening. Specialists currently disagree about how often women should get mammograms – too frequently and it drives up health-care costs with no benefit, not often enough and some early cancers might be missed. If the MIT system can learn to differentiate between people who will develop cancer within five years and those who won't, Yala says, it might allow physicians to personalize screening schedules and offer frequent mammograms only to those whose early scans show they are at high risk.

Researchers at Google are also trying to improve cancer screening. Medical groups in the United States and Canada already recommend screening certain people who are at high risk of lung cancer using computed tomography (CT) scans based on low-dose X-rays, and the same screening protocol is under consideration in the European Union. Computer scientists at Google wanted to see whether they could predict which people would go on to develop lung cancer by using AI to analyse low-dose CT scans of the lungs.

They collected about 43,000 scans from almost 15,000 people that had been amassed by the National Lung Screening Trial (NLST), a study run by the US National Cancer Institute. Of those, 638 people did not have cancer at the time of the initial scan, but were diagnosed within one year, with the cancer confirmed by biopsy⁴. "Our main goal was to try and predict whether someone ends up with lung cancer a year from when they got screened, or two years in some cases," says Shravya Shetty, a software engineer at Google in San Francisco, California.

In people with only one scan available, the AI outperformed all of the six radiologists who

also examined the CT scans to assess risk of lung cancer. The AI reduced the number of false positives by 11% and false negative by 5%. When there were two scans, the radiologists did about as well as the computer. Researchers hope that more accurate screening will lead to more effective treatment. "Ultimately what we want is patients to get their cancers caught earlier," says Daniel Tse, a medical doctor at Google Health who led the project.

The Google model is still very new, Tse says, and AI systems under development have a way to go before reaching widespread clinical use. "It does show great promise," he says, "but we're going to be doing further studies to see how the models interact in larger scales of data, new environments, things like that." The goal, he says, is to blend computer technology with the knowledge and skills of doctors, "and hopefully produce even better results than any one of the two could produce on its own".

Honing the tools

More training data and improved algorithms should increase the systems' accuracy. Medical data sets, even those that contain thousands of data points, are much smaller than the huge databases of online photos in which AI had its first big successes in image recognition. The FocalNet project, for example, had images from only 417 people to train on. But in these cases, scientists don't start from scratch. They use techniques and algorithms developed by other machine-learning researchers to jump-start their own models. They can also use AI to develop synthetic data sets – similar to the way in which some self-driving-car algorithms trained using data from video games such as *Grand Theft Auto*,

rather than from the real world.

Medical AI systems also need to be validated against populations other than the ones they were trained on; a system that seems to work on tests from one medical centre or on a particular population might fail when it's applied to a different group of people. The MIT team tested its model, which had been trained using data from a predominantly white population, to see if it worked equally well for black women – it did. It might be, Yala says, that the visible markers of breast cancer don't differ much between ethnic groups, but the only way to know that is to check. The team is testing its model, developed using data it collected in Boston, on populations from Detroit, Taiwan and Europe, and hope to do the same with data from Latin America. "We view it as our scientific responsibility to make sure that it works for everybody," Yala says.

Diversity isn't the only question to be explored. The approaches should work for other types of cancer as well. The Heidelberg researchers hope to publish a study this year on methylation profiling for sarcomas – cancers that develop in bones and soft tissues, which von Deimling says have a diagnostic error rate of 20% or higher. The researchers hope to move on from there to carcinomas, which develop in epithelial cells around organs and in the skin. Barzilay's group is investigating whether its system will work on pancreatic cancer; although there's no regular screening programme for pancreatic cancer, scans taken for other purposes might contain useful data (see page S12). And Tse says his group is looking at using AI to tell whether a skin lesion is cancerous.

None of the researchers expects AI to replace physicians, radiologists or pathologists. But with an ageing population, increased availability of diagnostic tests and growing emphasis on precision medicine, machine learning could help them to do their jobs by identifying the high-risk cases they should focus on and helping them to make decisions about uncertain diagnoses. Von Deimling doesn't imagine a computer will provide all the answers in medicine. "I would not leave diagnostics entirely to the classifier," he says. "This is not replacing a pathologist. It's just a tremendously powerful tool which should be in the hands of a pathologist."

Neil Savage is a science and technology journalist in Lowell, Massachusetts.

1. Capper, D. et al. *Nature* **555**, 469–474 (2018).
2. Cao, R. et al. *IEEE Trans. Med. Imaging* **38**, 2496–2506 (2019).
3. Yala, A. et al. *Radiology* **292**, 60–66 (2019).
4. Ardila, D. et al. *Nature Med.* **25**, 954–961 (2019).

ADAM YALA