



Computed tomography scan of a brain showing a haemorrhage (red region) that caused a stroke.

DIAGNOSTICS

# Tech support

*Artificial-intelligence technology could help radiologists and pathologists to diagnose disease.*

BY NEIL SAVAGE

The faster a person is diagnosed after a stroke, the sooner they can be treated and the more cognitive function can be preserved. But it takes time to identify the problem from a brain scan — an average of 87 minutes after the images have been captured, for cases flagged as urgent, according to one study<sup>1</sup>. In that time, tissue dies. Researchers have found that someone experiencing a stroke that cuts off flow through a large blood vessel will typically lose 120 million neurons, 830 billion synapse connections and 714 kilometres of nerve fibre in 1 hour — the equivalent of ageing by about 3.5 years<sup>2</sup>.

Cutting down the time it takes radiologists to diagnose people with stroke, therefore, could lead to better outcomes. “They notify me sooner and I can operate,” says Eric Oermann, a

neurosurgeon at Mount Sinai Health System in New York City. Oermann, who directs Mount Sinai’s artificial-intelligence (AI) consortium, AISINAI, has been studying whether the technology can help to speed up diagnosis.

He and his colleagues ran computed tomography (CT) images of brains through a type of AI known as a deep neural network<sup>1</sup>. First, the system was shown images annotated by radiologists and interpreted using a natural-language processing tool developed by Oermann and radiologist John Zech, now of Columbia University in New York City. Oermann’s hope was that, given enough example images to study, the algorithm would learn to identify features that distinguish a healthy brain from one experiencing an ischaemic stroke, a haemorrhagic stroke or the build-up of fluid known as hydrocephaly.

Their system was unable to match the

accuracy of a radiologist in diagnosing these three conditions. But it was good at quickly flagging brain scans that it deemed in need of attention — and therefore could be used to alert radiologists to examine urgent cases more rapidly. That, Oermann says, can make a big difference. Brain scans often sit waiting to be examined by radiologists for four hours or longer. If the system reshuffled the queue according to its assessment of urgency, the wait for a radiologist’s examination could shrink to just a few minutes for the most urgent cases, he says.

Oermann’s view is that AI will have a major impact on medical care. “I think everyone knows it’s going to radically change medicine in the twenty-first century,” he says. Indeed, researchers at hospitals, universities, small start-ups and computing giants are studying ways to use AI to identify and classify various conditions — from breast cancer and rare genetic disorders to depression. And although most algorithms with highly accurate diagnostic capabilities are years away from the clinic, a few might be much closer.

## FIELD OF VISION

The promise of AI lies in its powerful ability to identify patterns that can elude humans, either because the signs are too subtle or because they emerge only from huge sets of data. The technology has become more viable in the past decade through the use of deep neural networks, which are built to mimic the brain. In these, the nodes that make up the network represent individual neurons, connected by a mathematical weight that represents synapses. To train a neural network, researchers provide an input of example images, such as brain scans. This is translated into a set of numbers that might describe where each individual pixel in a scan falls on a 100-point scale from black to white. A hidden layer of nodes multiplies those input values by the weight of the connections and produces a numerical output. This is compared to the input, the weights are adjusted to make the two agree better, and the process is repeated. Eventually, the system develops a mathematical model of what a brain haemorrhage looks like, and can say how closely a new scan resembles the images on which it was trained.

Saeed Hassanpour, an electrical engineer who studies biomedical data science at Geisel School of Medicine at Dartmouth College in Hanover, New Hampshire, trained a neural network to classify a common type of cancer called adenocarcinoma from microscope slides containing samples of lung tissue<sup>3</sup>. How this condition is treated depends on the grade and stage of the tumour, but making that determination can be tricky. Affected cells can fall into any of five distinct subtypes, and most tumours contain a mixture. Some subtypes are associated with high survival rates, but if even a small number of cancer cells are of the deadlier variety, treatment must be more aggressive. Often, pathologists do not agree on what they’re seeing.

Hassanpour trained a neural network using a set of slides on which three Dartmouth pathologists had labelled the cellular patterns they saw. The neural network was then given a set of unannotated slides and asked to identify subtypes for itself. For any slide, it had a 66.6% chance of agreeing with at least one pathologist — slightly better than the 62.7% agreement between pathologists.

Such algorithms are unlikely to put humans out of a job, and that's not Hassanpour's aim. "I don't think it's going to be stand alone," he says. A system that just flags people who require special attention could be a powerful tool, says Stephen Yip, a neuropathologist at the University of British Columbia in Vancouver, Canada. In a set of 20 slides, an algorithm might identify 3 that contain an abnormality needing further examination. The pathologist could concentrate on these flagged slides, leaving the images less likely to show signs of a problem until later. "I tell people that machine learning is not going to replace their job, but it's going to make their job easier," Yip says.

### EARLY OPPORTUNITIES

One reason neural networks can't yet outperform humans is that the models are trained to interpret just one type of information, whereas physicians consider many factors when they make a diagnosis. "Almost all diagnoses in medicine are done using all the information at hand, which is rarely to never a single test," Oermann says. Knowing whether a person came in with headaches or as a result of a motor vehicle accident, for instance, will lead a radiologist to look at a brain scan differently.

Truly powerful automated diagnostic systems will come when researchers integrate multiple data types into the same model, says Yip. This could include not only pathology and radiology, but also genomics, electronic health records and even lifestyle data. "You really need multiple levels of data for us to discover something that's truly amazing," he says.

In the few cases in which a single test can discover the presence of a disease, AI-based diagnostics might be just around the corner. Researchers at Google, for instance, are training computers to identify the eye disease diabetic retinopathy from images of people's retinas, as do ophthalmologists. And Oermann says that, for diagnoses that are based solely on examining samples on slides — such as identifying cancerous tissue — using AI to make the call "is completely realistic."

Pathologist Andrew Beck agrees. He left his faculty job at Harvard Medical School to co-found PathAI, a start-up in Boston, Massachusetts, that is trying to automate diagnosis. "The images themselves just contain so much data," he says, and neural networks are becoming ever-better at ferreting out all the information. He works with pharmaceutical companies that are developing immunotherapies to target programmed death-ligand 1 (PD-L1), a protein that lets cancer cells evade

```

268 num_removed=0
269 for i in range(0,df_data.shape[0]):
270     if df_data['Accession Number'].iloc[i] in labeled_cases and df_data['Accession Number'].iloc[i] not i
271         train_data_lda[i]=False
272         labeled_indices.append(i)
273         num_removed += 1
274     else:
275         not_labeled_indices.append(i)
276 unlabeled_corpus = [corpus[i] for i in not_labeled_indices]
277 labeled_corpus = [corpus[i] for i in labeled_indices]
278
279 return corpus, unlabeled_corpus, labeled_corpus, dictionary, labeled_indices
280
281
282 def build_d2v_corpora(df_data,d2v_inputs,labeled_indices):
283     """
284     Build corpora in format for doc2vec gensim implementation
285     Args:
286     df_data: a dataframe consisting of the original set of excel files with report text and accession id
287     d2v_inputs: list of lists of tokens, where each entry in d2v_inputs corresponds to a report
288     labeled_indices: indices of labeled reports (and those we treat as labeled due to TRAIN_INDEX_OVERRID
289     Returns:
290     unlabeled_corpus: a corpus consisting of unlabelled texts that will be used for feature construction
291     labeled_corpus: a corpus consisting of labelled held-out texts that will be used for Lasso regression
292     total_unlabeled_words: count of total words in unlabeled corpus
293     """
294
295     SentimentDocument = namedtuple('SentimentDocument', 'words tags')
296     unlabeled_docs = []
297     labeled_docs = []
298     total_unlabeled_words=0
299     i=0
300     for line in d2v_inputs:

```

John Zech, now at Columbia University, New York City, helped to develop code for the AISINAI project.

the immune system. Pathologists stain tissue and then count individual cells that express PD-L1, but an AI system should be able to do that with great precision, as well as capturing other information a human would not, such as the spatial relationships between cells. The system could then use those data to predict whether a particular drug would be effective for an individual patient. "It really is a very different type of data that we can now extract out of these images," Beck says.

### HUNGRY FOR DATA

One factor standing in the way of using deep learning in medicine is the dearth of data sets of sufficient size to train and test the models. Part of the reason that AI became so good at recognizing objects in photos is that the internet is awash with images. ImageNet, a tool that is used to train many neural networks, contains upwards of 14 million photos, all of them

**"I tell people that machine learning is not going to replace their job, but it's going to make their job easier."**

labelled by people to tell the machine what they show. Data sets from medical imaging are growing, but are a fraction of the size. Oermann and his team trained their system on 96,000 cranial CT scans. But for his model to outperform humans, he says, he would have needed perhaps ten times as many. In pathology, even fewer images are available, Hassanpour says. Digitizing slides has long been expensive and time-consuming, although hospitals are doing more of it as costs fall and technology improves.

Another problem is labelling the images. ImageNet developed its labels using the Amazon Mechanical Turk crowdsourcing platform to hire people to circle cats, lamp posts and noses in photos at a cost of pennies per image. But circling tumours or lymphocytes in medical images requires trained

specialists. One way to accelerate this process in medical imaging is an approach called active learning. In this method, the computer is shown weakly labelled images that state only whether an image contains a haemorrhage or not, for instance. The algorithm then attempts to label the affected areas in each image. Specialists add labels where the computer came up short and remove those placed in error, and these images are used for a new training round. After a few repeats, the result is a well-labelled set of images.

Oermann is also experimenting with using synthetic data for initial training of his system, which can then be refined using real medical data. Scientists developing algorithms for self-driving cars have taken a similar tack, training their models on the video game *Grand Theft Auto* before showing them actual roads<sup>4</sup>.

Before AI systems can be used in medicine, they will have to go through clinical trials to establish their validity. "Health care in general is very sensitive, and we need to make sure that the model actually provides the value and benefit to clinicians and, more importantly, the patient," Hassanpour says. But the need to lessen the diagnostic burden on pathologists and clinicians is clear. Advances in precision medicine, Beck says, will bring more diagnostic tests and more treatment options, and will only increase the workload. "Demand's going to go up, not down," he says.

For now, however, AI systems will be there to help doctors to make a diagnosis, not to put them out of a job. "There's a lot of human factors that go into health care," Yip says. "Nothing can replace the human touch." ■

Neil Savage is a freelance writer based in Lowell, Massachusetts.

1. Titano, J. J. *et al. Nature Med.* **24**, 1337–1341 (2018).
2. Saver, J. L. *Stroke* **37**, 263–266 (2006).
3. Wei, J. W. *et al. Sci. Rep.* **9**, 3358 (2019).
4. Martinez, M. *et al. Preprint at* <https://arxiv.org/abs/1712.01397> (2017).