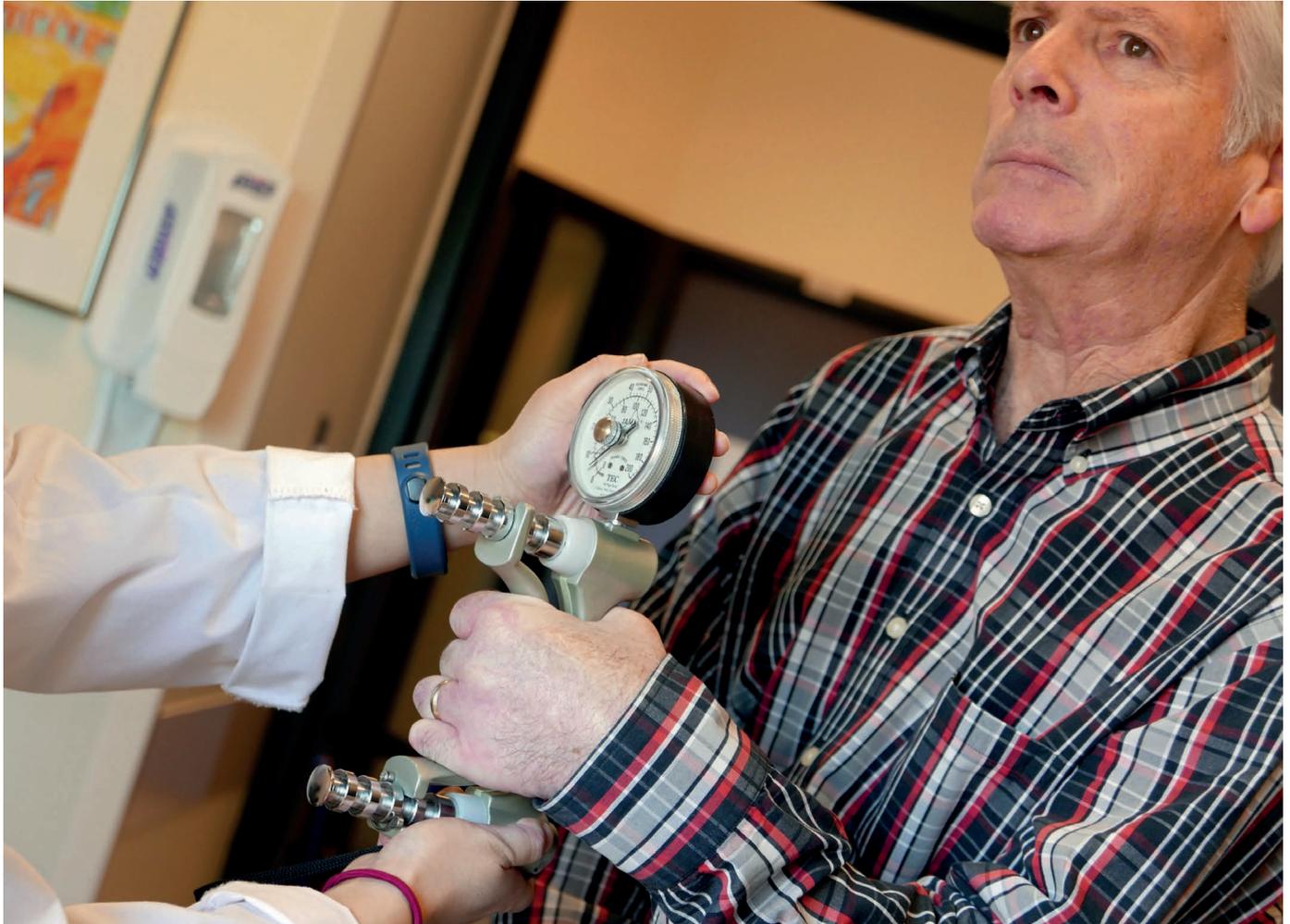


MACHINE LEARNING

Calculating disease

Machine learning might identify patients earlier, predict their outcomes better, and assign them more efficiently to appropriate clinical trials.



MILAN KAROL/ROBERT PACKARD CENTER FOR ALS RES./JOHNS HOPKINS

A patient enrolls in the Answer ALS project, which aims to improve characterization of ALS subtypes.

BY NEIL SAVAGE

For neurologist Peter Bede, seeing someone he thinks might have amyotrophic lateral sclerosis (ALS) can be vexing. Unlike many disorders, even other neurological diseases, ALS is not an illness that can be identified by an examination and a few lab tests.

“If I suspect multiple sclerosis, I do an MRI and do a lumbar puncture and that pretty much will confirm diagnosis,” says Bede, an ALS researcher at Trinity College Dublin. “When I do an ALS assessment, I don’t have these markers.”

Instead, he’ll examine the patients, then tell them to return a month later to see whether the symptoms have worsened. Only after watching

the deterioration continue over seven or eight months will he be able to say for sure that it’s ALS. Instead of relying on a blood test or X-ray to confirm his suspicions, Bede has to let time pass and note changes in symptoms. “That’s incredibly frustrating for me as a physician, because I suspect it, but I can’t be 100% sure until I see the patient a few more times,” he says.

It’s also frustrating for patients, who on average spend 12 months between the onset of symptoms and a conclusive diagnosis. During that time, they are not given one of the two drugs approved to slow the disease, and they often undergo unnecessary surgery because their muscular weakness is misdiagnosed as a compressed disc or carpal tunnel syndrome. The delay slows recruitment into clinical trials and confounds comparisons of drug candidates.

ALS is a challenging disease for researchers because it is so hard to predict its course. Some people die quickly, whereas others can survive for decades. But now, machine-learning systems are providing researchers with new insights into ALS. Scientists are trying to diagnose the condition earlier by applying pattern-recognition algorithms to magnetic resonance imaging (MRI) scans of the brain and spinal column. They are combining data gleaned from gene sequencing with clinical findings in the hope of dividing patients into slow, medium, and fast progressors — the better to assign them to more precisely targeted clinical trials. They’re applying acoustic analysis to speech in hopes of making an earlier diagnosis and using natural-language processing to search for signs



The Answer ALS app's 'trace task' tests fine motor control.

of neural degeneration. And they are combing through electronic medical records to see if they can identify patients years before the onset of conventional ALS symptoms. By combining these advanced technologies, clinicians and researchers are trying to create an early-warning system for this devastating condition.

BIOPSY BY NUMBERS

Bede is developing an algorithm to perform what he calls virtual biopsies on patients' brains¹. He trained a set of algorithms on 110 MRI scans, half of them from people with ALS and half from healthy controls, telling the system which was which. The algorithms, which map out the scans in three dimensions and apply various statistical analyses, compared certain characteristics of the brain — the thickness of white matter, the volume of basal ganglia — and learnt how they differed between healthy people and people with the disease. Bede then showed the system 40 more scans. The computer correctly identified 90% of the people with ALS.

Bede has also trained the computer to predict survival using MRI images in conjunction with clinical data². These clinical data included age at symptom onset, the area of the body where the disease first manifested, and physical disability measured by the standard ALS Functional Rating Scale. Clinical data on their own predicted which patients would be alive after 18 months with 67% accuracy. Using MRI data alone gave an accuracy of 77%. Combining the two, the accuracy climbed to 79% and the rate of false negatives declined.

Although encouraged by the results, Bede would like to use a much larger data set to validate the conclusions and refine the algorithm. And although he has shown that the computer can distinguish between a healthy person and a person with ALS, even better would be the ability to separate ALS patients from those with neurological disorders that can mimic the disease, such as frontotemporal dementia.

To provide larger data sets, Martin Turner,

a neurologist at the University of Oxford, UK, helped to form the Neuroimaging Society in ALS (NiSALS), which is collecting MRI scans of patients and storing them in a repository in Jena, Germany. So far, Turner says, NiSALS has between 500 and 600 patient records.

Whereas Bede looks for structural changes in the brain as a marker of ALS, Turner has a different goal: to determine how quickly the disease is progressing. He uses functional MRI scans to look at the patient's blood oxygen level, which indicates which parts of the brain are active during a given activity. When a person in the scanner isn't doing or thinking of anything in particular, the blood oxygen level fluctuates in specific patterns in different regions of the brain. Neurologists have identified ten sets of patterns that look the same in every individual.

But in people with neurological conditions such as ALS, the patterns start to degrade in a way that provides a fingerprint of the disease. "You couldn't possibly make anything of that just visually," Turner says. The computer is able to find patterns that distinguish fast-progressing patients from slow ones. "It's a much more objective biomarker, which you could only pull out with a machine."

Researchers have noted that levels of some proteins in blood and cerebrospinal fluid are different in people with ALS and in those without the disease. It might be possible eventually, Turner says, to detect patterns in how those levels change over time, and combine them with the functional MRI patterns to refine the classification.

Determining how fast an ALS patient is likely to progress is important to clinical-trial design. If a drug candidate is supposed to slow progression of the disease, and many of the participants in a trial have disease that is already progressing slowly, any effect may be hard to see. The current standard in ALS trials is to give half the group the candidate drug and half a placebo, and then see who is still alive after 18 months.

"That's a very blunt tool, and it takes a long time," Turner says. "What you really want to be able to do is measure a bunch of markers — by a mix of imaging and neurochemistry — in patients now, and then again in six months, and have that be enough to see a clear difference in the disease signature."

Another approach to sorting patients by how fast their disease progresses uses a computing approach called deep learning, the same technique that has vastly improved Google's ability to recognize pictures of, say, cats. Deep learning uses an artificial neural network, in which simple silicon processing units, called artificial neurons, are connected to many others in a complex pattern, like neurons in the brain. Data are fed into the first layer of artificial neurons, each of which performs a small calculation and then sends the results on to many others, which perform further calculations and pass those results to another set, creating a complex set of inputs and outputs until a single result finally comes out at the other end. The computer could, for example, pick out 50 or 60 structural features from an MRI scan, then double or triple that number in the 'hidden' layers of the network before classifying the ALS patient as likely to be a short-term, medium-term or long-term survivor³. There could be hundreds of interactions going on between individual processors in the hidden layers, says Martijn van den Heuvel, a computer scientist at the University Medical Center Utrecht in the Netherlands.

The sheer volume of calculations makes the neural-network algorithm much more powerful than older machine-learning techniques. "Even with a relatively small data set you can still make good predictions," van den Heuvel says. But it requires a lot of computing power. With today's computers, processing an MRI image into data the computer can use takes an hour; it then takes many days of computation to extract the salient features from the data and give a prognosis for each patient.

VIRTUAL PATIENTS

MRI scans are not the only way to predict how fast a person's ALS will progress. Clinical trials collect a lot of data, and computer scientists are examining ways to build predictive models from those. For instance, Origent Data Sciences, a biotech start-up in Vienna, Virginia, has developed three computer models using clinical information. One relies on patients' ALS Functional Rating Score to predict which category they fall into. A second model relies on vital capacity, the maximum amount of air the lungs can expel. The third looks at how long people actually lived; the computer then compares a variety of clinical measurements from new patients to those from the model to predict how long the new patients will survive.

The hope, says Dave Ennist, the company's chief science officer, is to combine the models

into a system that predicts outcomes for patients in a given group, then use that aggregate model as a virtual control arm in clinical trials. Instead of comparing people given a drug candidate with those taking a placebo, the trial would compare them with a computer model of how they would have fared without the drug. Origent has a US\$500,000 grant from the ALS Association in Washington DC to test its approach using data from an ongoing trial of the drug tirasemtiv. This drug is being developed by Cytokinetics, a biotechnology company in South San Francisco, California. The goal is to show that the computer can accurately predict the outcome of the clinical trial.

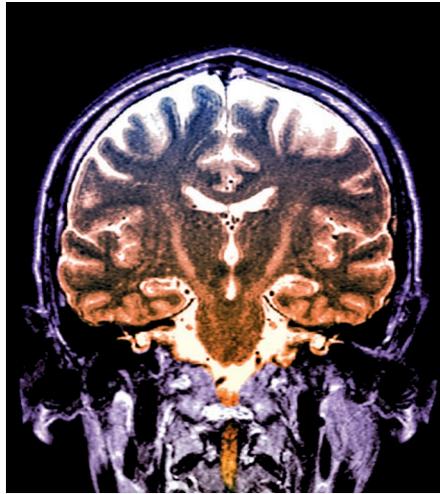
Aside from stratifying patients by their expected progression rate, clinical researchers would also like to know whether there are different subtypes of ALS, caused by different molecular pathways that might provide targets for drugs. That's the main question being asked by Answer ALS, a project that's collecting an array of data from 1,000 patients across the United States. In addition to detailed clinical data, researchers are turning blood samples into induced pluripotent stem cells, and using those to make motor neurons on which they can run tests. They're also determining patients' proteomes, epigenomes and transcriptomes, among other omes.

All told, the project will generate about 6 billion data points per person. "You apply big-data approaches to that, then you can hopefully start to see the patterns about different subtypes of the disease," says Ernest Fraenkel, a computational biologist at the Massachusetts Institute of Technology in Cambridge, who is in charge of computing for Answer ALS. "Then you can say, 'OK, how do we fix this?'"

THE POWER OF SPEECH

Answer ALS is also developing a smartphone app to collect more clinical data than can be gained from an office visit every few months. The goal is to track disease progression in something closer to real time. For instance, the app will ask patients to take a deep breath and speak as many numbers as they can, which provides a crude measure of lung function and, over time, can show how fast their capacity is diminishing. As a way to measure muscular coordination, people will be asked to trace an image on the phone's screen. And in one experiment aimed at detecting signs of mild cognitive impairment, they'll be shown a drawing of an ordinary scene — a person washing dishes, for instance, or a child climbing a tree — and asked to describe it. The descriptions will be evaluated by Watson, IBM's artificial-intelligence system. IBM hopes that by analysing details such as whether patients use simple or complex words and sentences, the algorithm will detect early cognitive changes.

Speech might also provide a means of diagnosing ALS earlier, because changes in motor control affect movement of the lips,



MRI scans help predict how fast ALS will progress.

tongue and vocal cords. Jun Wang, a computer scientist who directs the Speech Disorders and Technology Laboratory at the University of Texas, Dallas, asked a group of people with ALS and a group of healthy controls to speak a series of sentences while a computer extracted hundreds of acoustic features, such as how much the pitch of the voice varied, and measures of vocal-cord stability called shimmer and jitter. Wang then applied machine learning to the data to see if the computer could distinguish people with ALS from healthy participants.

Standard machine learning identified 65% of the people correctly. When Wang used deep learning instead, accuracy shot up to 92% in the small group of patients that he tested. Adding measurements of lip motion — which can be taken with a smartphone camera — improved the accuracy further. Wang hopes to have a smartphone app ready by the end of 2017 so he can collect data from a much larger number of people to validate the technique.

COMBING THE RECORDS

Cutting down the year-long delay between first symptoms of ALS and a definitive diagnosis would certainly help in leading patients to therapies and enrolling them in clinical trials. But identifying people with ALS even earlier could advance both the understanding and the treatment of the disease. "The earlier you can start studying these patients, the more clues you might have about what causes the disease," says Nazem Atassi, a neurologist at Massachusetts General Hospital and Harvard Medical School in Boston. "The earlier you start therapies, the more likely you have success in delaying disease progression."

With funding from Mitsubishi Tanabe

Pharma of Japan, which in May won regulatory approval for the first new ALS drug in 20 years, and help from HVH Patient Precision Analytics, a medical-analytics company based in New York City, Atassi set out to see if he could spot ALS patients years before their diagnosis. He looked at a commercially available database containing insurance claims for more than 170 million Americans¹. Using analytical techniques developed by researchers who work with big data, he found nearly 14,000 people who, according to the diagnostic codes used in billing, had ALS and had been seen by doctors for up to 5 years before their ALS diagnosis between 2010 and 2016.

On the basis of the tests and treatments that they had been given, the people with ALS stood out from the crowd. "These patients are different from the general population going five years back before ALS diagnosis," Atassi says. For instance, people who were eventually diagnosed with ALS were evaluated for non-specific nervous-system disorders more often than the general population. Multiple sclerosis was one of the top diagnoses five years before the ALS diagnosis, but dropped off later. At three years before diagnosis, malaise, fatigue and gastrointestinal disorders were among the top ten diagnoses and were more common than in the general population; they became more common over time. Skin disorders were more common in the ALS group at five years out, then dropped off. Atassi doesn't know why, but speculates that muscle twitching was misread as itching.

If he can refine the algorithm, it might be used to flag up people who have a particular pattern of symptoms at a particular time, and prompt a physician to order certain tests or to at least keep an eye on the patient for further signs of ALS.

Identifying people with ALS five years earlier than would otherwise happen might allow researchers to study their biology and see how the development of the disease progresses, possibly providing clues for future therapies. Even without that much early warning, the researchers are confident that machine learning will help them push diagnosis back, and get patients into treatment and clinical trials earlier. Earlier diagnosis may eventually lead to new pharmaceuticals, and in the short term could improve quality of life. "Even if we reduce the time to diagnosis by three months," Atassi says, "that is a big advance." ■

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

1. Bede, P., Iyer, P. M., Finegan, E., Omer, T. & Hardiman, O. *NeuroImage Clin.* **15**, 653–658 (2017).
2. Schuster, C., Hardiman, O. & Bede, P. *PLoS ONE* **11**, e0167331 (2016).
3. van der Burgh, H. K. *et al. NeuroImage Clin.* **13**, 361–369 (2017).
4. Grawbowski, T. *et al. Neurology* **88**, Suppl. P4.119 (2017).