WHITE PAPER

DELL EMC AI RESEARCHERS ACHIEVE DIAGNOSTIC ACCURACY RATES THAT MATCH HUMAN RADIOLOGISTS

COLLABORATIVE EFFORT USES MACHINE LEARNING, PUBLIC DATA, OPEN SOURCE TOOLS TO ACHIEVE BREAKTHROUGH

With healthcare costs soaring and the shortage of physicians expected to reach historic highs over the next decade, artificial intelligence is increasingly being seen as the best way to stem a vicious cycle of spiraling costs.

More than 300 healthcare-related startups have emerged since 2016, and the pace of innovation is increasing as technology costs come down and software used for machine and deep learning becomes more capable and easier to access.

AI stands to transform the healthcare world as computers take on tasks that were once limited only to human experts. Computers are already far better at general-purpose image recognition than humans, and they are approaching parity in areas like voice recognition and language translation. But recognizing faces in a crowd isn't the same as identifying the warning signs of a tumor or heart defect, a discipline that requires years of specialized training.

AI is catching up, though, and thanks to collaboration in the medical community, it's beginning to augment even the most seasoned physicians in diagnostic scenarios.

SPONSORED CONTENT

FROM IDG

IDG Communications, Inc.

A team of data scientists at Dell EMC, with support from Intel, recently demonstrated the dramatic potential of deep learning algorithms to assist professional radiologists in the detection of pneumonia and emphysema, conditions that kill about 60,000 people in the United States alone each year. The team built upon a neural network developed at Stanford University called CheXNet that scans chest x-rays and calculates the probability that pneumonia is present.

Tests showed that the model the Dell EMC team built not only performed better than the original CheXNet model, but also outperformed baseline tests in 10 out of 14 different categories, including diagnosis of emphysema, a lung condition that afflicts an estimated 3.5 million Americans.

LOW-COST INNOVATION

The achievements of the Dell EMC team are significant not only for their clinical value, but also for the cost-effectiveness of the solution. The team used off-the-shelf hardware components in a "scale-out" configuration and trained the model using software that's freely available under open source licenses.

Their platform was the Zenith cluster, a system built with the Intel Scalable Systems Framework that works without the specialized graphics processing units (GPUs) that can significantly increase the cost of AI projects. While GPU-enabled computers are popular with AI researchers for their ability to split up large tasks and process them in parallel, they are also significantly more expensive and can require more bandwidth. The Dell EMC team was able to achieve comparable performance at a fraction of the hardware and software costs.

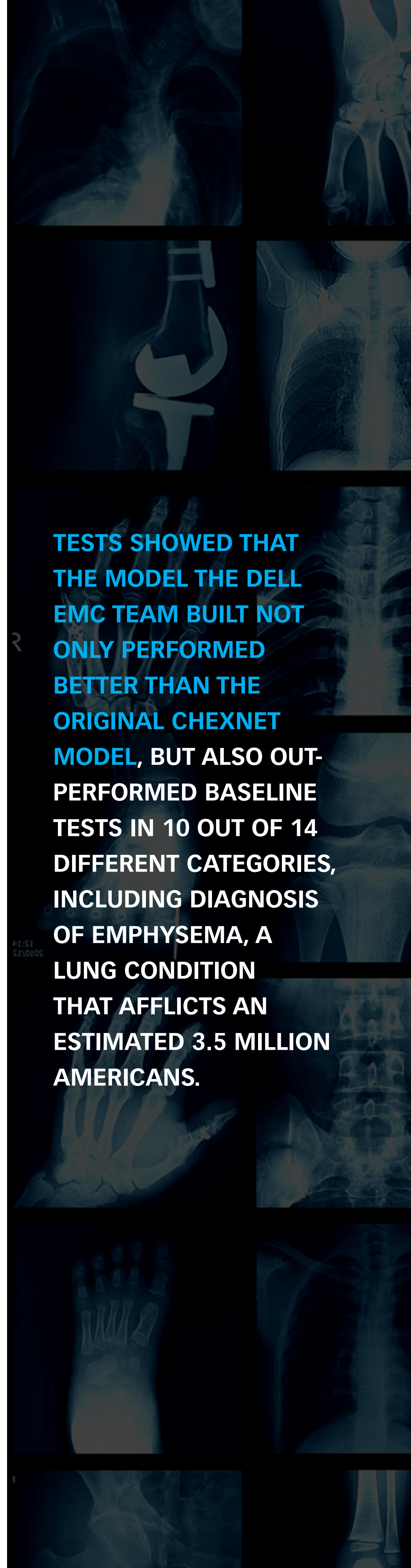
CheXNet demonstrates the power of technology combined with collaboration. In the fall of 2017, the National Institutes of Health Clinical Center released more than 100,000 anonymized chest x-ray images and their corresponding data to the scientific community. It then challenged data scientists to develop AI solutions that could analyze the images as effectively as trained radiologists.

Dell EMC's AI team spotted an opportunity to not only advance the medical community's body of knowledge, but to prove that practical AI applications could be run without the use of specialized hardware. Universities and healthcare organizations tend not to share the data they gather publicly for compliance and competitive reasons, so the sudden availability of so much source material was an unusual event, says Lucas Wilson, the Dell EMC AI research scientist who headed up the effort.

The problem set was also interesting. Diagnosing x-rays is difficult even for radiologists with years of experience. The grayscale images that x-ray machines produce don't provide much detail beyond the density of a mass, and pneumonia is a particularly tricky condition to diagnose because it can mimic many other benign abnormalities. Spotting it correctly requires knowledge of anatomy, physiology and pathology, not to mention years of experience.

The image-recognition task is an even tougher problem for machines. While computers have achieved proficiency in identifying people and animals, much of that task is a matter of matching a few basic patterns, since shapes don't vary much from subject to subject. However, finding pulmonary masses and cancerous tumors is different, since the objects don't adhere to any specific pattern. That's what makes radiology "as much art as science," Wilson says.

The Stanford team created a baseline by culling a random set of 420 chest x-rays and submitting them to four experienced radiologists, who provided their own diagnoses.



TESTS SHOWED THAT THE MODEL THE DELL EMC TEAM BUILT NOT ONLY PERFORMED BETTER THAN THE ORIGINAL CHEXNET MODEL, BUT ALSO OUT-PERFORMED BASELINE TESTS IN 10 OUT OF 14 DIFFERENT CATEGORIES, INCLUDING DIAGNOSIS OF EMPHYSEMA, A LUNG CONDITION THAT AFFLICTS AN ESTIMATED 3.5 MILLION AMERICANS.

Their opinions were used as a baseline for training on a larger set of 10,000 examples. CheXNet was able to better identify pneumonia.

HOW MACHINE LEARNING IS DIFFERENT

Training is a unique feature of machine learning that distinguishes it from other types of data processing. Traditional applications take a set of inputs and apply operations that result in outputs, such as calculating an electric bill based upon a customer's power consumption. When all is working properly, the same set of input data should always produce the same output.

Machine learning doesn't work that way. Instead, it starts with a baseline of facts and sorts through the data repeatedly, looking for patterns that correlate with what it already knows. It then adds that new information to its knowledge base, growing smarter with each pass through the data. For example, a machine learning program that knows that cats have whiskers and pointy ears can repeatedly sort through a database of images to figure out that they also have claws, soft fur and long, flexible tails.

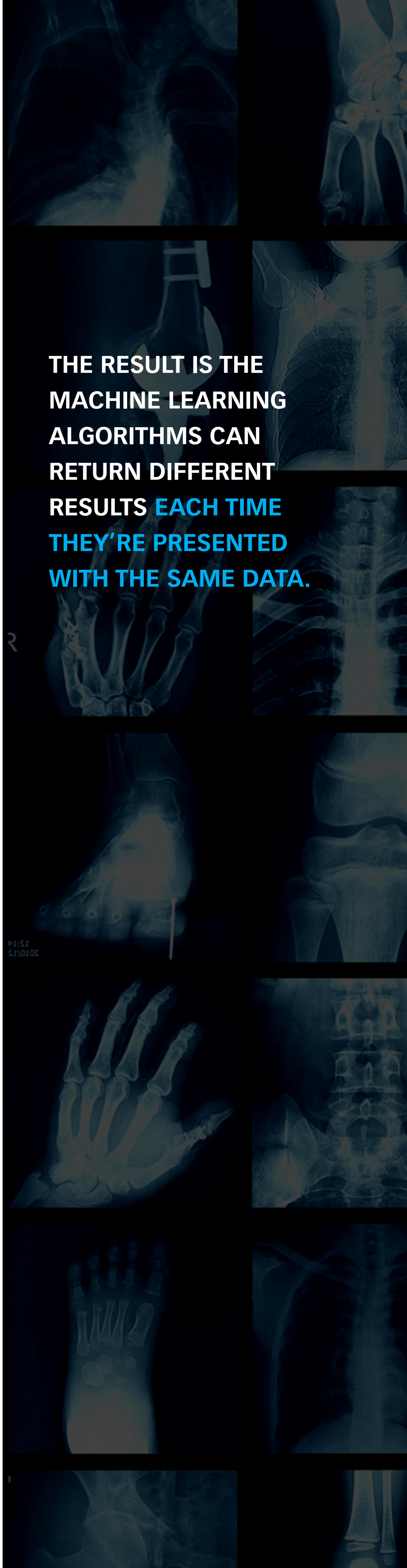
The result is the machine learning algorithms can return different results each time they're presented with the same data. That's because they're designed to continually iterate through the information and spot patterns, which may differ from case to case. This distinctive characteristic makes testing difficult because the precise outcomes aren't known. That's why the baseline of expert opinion was so important to measuring results.

The Dell EMC AI Engineering team used a "convolutional neural network (CNN)," which is a type of artificial intelligence that works a lot like the human brain. The network splits the images and instructions up among many individual processing nodes, each of which passes its findings to a control layer. The results are aggregated at that level and fed back to the processing nodes as they prepare for another pass through the data. One of the advantages of convolutional neural networks is that they can learn without having to have a lot of rules predefined by humans. They can also operate on very large data sets because processing is divided up among many nodes, each of which works on a subset of the problem.

But CNNs also require a lot of computer power. That's why they're typically built with a combination of CPUs and accelerators working in parallel. By processing data in a distributed model, organizations can leverage off-the-shelf components to create supercomputer-like performance. That's paying big dividends in the data-intensive healthcare field, enabling physicians and scientists to glean insights from volumes of data that are far too large for a human to process.

For example, the cost of sequencing a human genome has dropped from $100 million in 2001 to less than $1,000 today, making it arguably the greatest deflation in history. However, analyzing the roughly 4 TB of data contained in a single genome is a massive computing problem. Organizations like the Translational Genomics Research Institute are now able to unlock the value of genomes using standards-based infrastructure from Dell EMC that has cut the time needed to process a single genome from weeks to hours.

Cost efficiency was on the minds of the Dell EMC AI Engineering team for practical reasons. The company was preparing to roll out a version of its Ready Solutions for AI, incorporating powerful servers, storage and a set of AI frameworks and libraries built for deep learning workloads. The team wanted to see what it could accomplish without adding expensive GPUs. Using a system with a single Intel® Xeon® Scalable CPU, they were able to achieve about 80% accuracy, slightly below that of human radiologists. But the precision came at the expense of time. It took several days to complete the training of a single model.



SCALING THE SOLUTION

So, the researchers brought more servers into the fold, tying them together in a “scale out” architecture that achieves almost linear performance improvements by adding more machines to the cluster. That enabled them to dramatically increase the number of images being processed in parallel from eight to as many as 32,000.

The computer was able to process more images, but not to run the necessary classification algorithms to understand what it was seeing. Dr. Wilson compared the process to cramming for a test: a student may absorb a large amount of information in a short time but will have trouble contextualizing or committing it to memory.

Leveraging the Intel® Scalable Systems Framework, the team constructed a net made up of more than 200 Intel® Xeon® CPUs operating in parallel. They also experimented with several popular neural network designs used in image processing while varying batch sizes to achieve the optimal combination of performance and accuracy.

The result was an astonishing 1,700-fold increase in speed, which reduced training time from the original five days to just 15 minutes. Best of all, accuracy actually improved; the new model successfully diagnosed pneumonia and emphysema 90% of the time, better than the performance of professional radiologists.

This study achieved several important milestones. It proved that large-scale parallel processing for AI applications could be conducted without the use of specialized hardware. “We wanted to demonstrate that computational problems don’t necessarily require a GPU,” Wilson says. “People use their CPUs for multiple things, and already they have them in their data center.” That makes AI more accessible to more people.

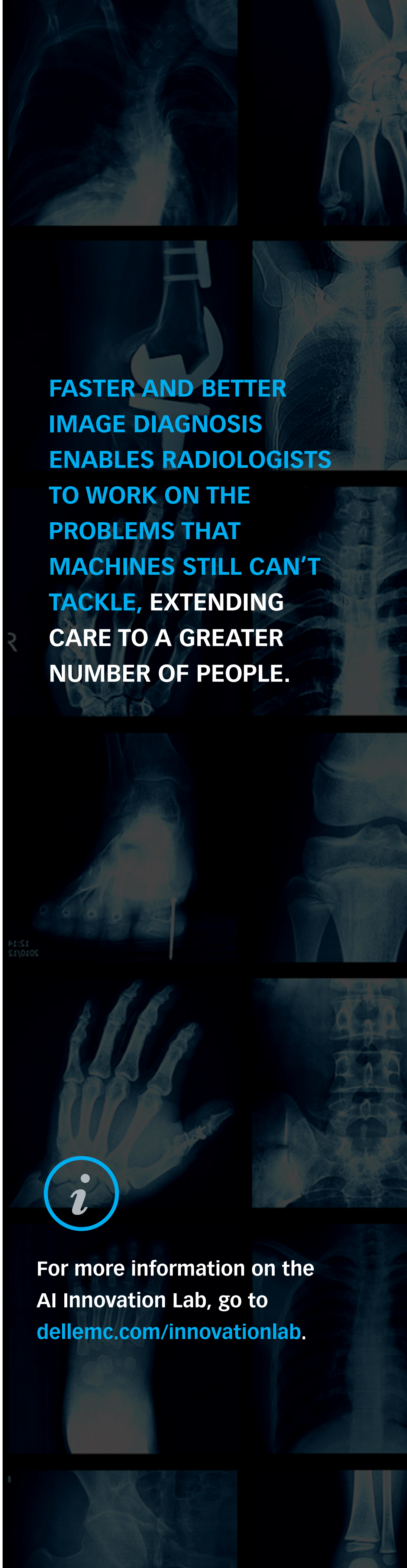
The project also proved that large-scale image processing could be done at a speed that enables clinicians to act on the spot, potentially saving lives. University-led AI projects are typically more concerned with results than with speed, Wilson says. The Dell EMC data science team proved that it was possible to have both.

Lower costs also improve accessibility. Many poor and remote areas lack sophisticated medical equipment, but x-ray machines are relatively available. Machine learning algorithms make it possible for healthcare professionals in those areas to have the benefit of higher-quality diagnosis with hopefully less wait-time.

The team also advanced the state-of-the-art in image diagnostics to new levels, an important contribution to the community of deep learning and machine learning professionals, who build upon each other’s accomplishments. The collaborative nature of AI research means that developers don’t have to start from scratch each time. They can take the research that taught machines to recognize cats and extended into new dimensions. “The greatest opportunity for advancement is taking work somebody else has already done and improving upon it,” Wilson said.

Ultimately, it’s about technology enhancing human performance. While some people debate whether AI-powered machines will eliminate jobs, professionals in the chronically short-handed healthcare field welcome the assistance. Faster and better image diagnosis enables radiologists to work on the problems that machines still can’t tackle, extending care to a greater number of people.

In the end, that’s what really matters.



FASTER AND BETTER IMAGE DIAGNOSIS ENABLES RADIOLOGISTS TO WORK ON THE PROBLEMS THAT MACHINES STILL CAN’T TACKLE, EXTENDING CARE TO A GREATER NUMBER OF PEOPLE.

For more information on the AI Innovation Lab, go to dellemc.com/innovationlab.