# AI and Machine Learning in Medical Imaging: From Potential to Reality

Powerful computing and storage capabilities required to move forward

## Introduction

The potential for artificial intelligence (AI) and machine learning in the healthcare industry is considerable. However, realizing this potential is not guaranteed. The challenge: While AI and machine learning offer a bevy of prospective benefits, software developers and researchers need to implement and then routinely rely upon a technology infrastructure that offers both the processing and storage capacity that will enable them to efficiently design machine learning solutions. By doing so, they can then bring never-before-imagined quality and efficiency improvements to the everyday delivery of care.

## AI and Machine Learning: Vast Potential

AI and machine learning carry great promise for the healthcare industry, according to experts who testified before the Senate Science, Commerce and Transportation Committee's Subcommittee on Communications, Technology, Innovation and the Internet in December of 2017. For example, Victoria Espinel, president and CEO of BSA|The Software Alliance, pointed to a 2016 Frost & Sullivan report that predicts AI has the potential

to improve health outcomes by 30% to 40%. She specifically pointed out that AI is leading to improved outcomes, not by replacing the decision-making of healthcare professionals, but by giving them new insights into vast amounts of data.

"AI tools are powering machine-assisted diagnosis, and surgical applications are being used to improve treatment options and outcomes," testified Espinel. "Image recognition algorithms are helping pathologists more effectively interpret patient data, thereby helping physicians form a better picture of patients' prognosis. The ability of AI to process and find patterns in vast amounts of data from disparate sources is also driving important progress in biomedical and epidemiological research."[1]

## Trailblazing Organizations Already Tapping into AI and Machine Learning

Indeed, many trailblazing organizations are discovering that AI and machine learning carry significant potential in the medial imaging realm. Consider the following:

- Mayo Clinic neuroradiologists are using AI to find genetic markers in MRI scans.[1] Instead of testing samples collected during surgery, AI leverages MRIs to predict brain tumor genomics. The work, called radiogenomics, "reflects the almost unthinkable thought that in the appearance of images, we can figure out the genomic properties of tumors," said Bradley Erickson, MD, a Mayo Clinic neuroradiologist.[2]
- Stanford researchers have developed an AI algorithm that can diagnose up to 14 types of medical conditions and is able to diagnose pneumonia from medical images.[2]
- "Interpreting X-ray images to diagnose pathologies like pneumonia is very challenging, and we know that there's a lot of variability in the diagnoses radiologists arrive at," said Pranav Rajpurkar, a graduate student in the Machine Learning Group at Stanford and co-lead author of a paper about the CheXNet application. "We became interested in developing machine learning algorithms that could learn from hundreds of thousands of chest X-ray diagnoses and make accurate diagnoses."[3]
- Memorial Sloan Kettering Cancer Center is collaborating with an imaging company to improve the diagnosis of prostate cancer. The healthcare provider is using the company's investigational imaging platform to retrospectively study the MRI images of 200 patients who were diagnosed with prostate cancer and underwent radical prostatectomy. The study is examining the use of automated texture analysis and machine learning to classify lesions on MRI images to determine whether the lesions are aggressive forms of the disease. The collaboration could be useful "in exploring potential applications of machine learning for the detection and classification of disease, with a particular focus on prostate cancer," said Hedvig Hricak, MD, chair of the Department of Radiology at Memorial Sloan Kettering. "Machine learning and artificial intelligence could ultimately provide precise tools that will augment our ability not only to diagnose prostate cancer but to potentially characterize it biologically and predict its behavior."[4]

## Machine Learning: Helping with Rising Radiology Workload

In addition to these advances in clinical quality, machine learning can help radiologists can gain the efficiency needed to cope with rising workloads, according to Esteban Rubens, global enterprise imaging principal at Pure Storage.

"In the United States, on average, radiologists have one point five seconds to look at each image and it's getting worse. So, some kind of machine assistance will essentially be mandatory," Rubens said. "Radiologists don't really have time to manually look at all the images generated. AI and machine learning can be a big help as these technologies can separate the wheat from the chaff and identify all the normal images that the radiologists don't need to look at."

In fact, as machine learning applications become more common, all images will be evaluated by a machine first to make some determination of normal or abnormal. Radiologists will then have more time to spend with the images that truly requires their advanced expertise.

## Support System Required to Move Forward

The challenge for researchers looking to develop medical imaging machine learning solutions rests in creating a technology infrastructure that can support these data intensive initiatives.

"The common denominator is you need a huge amount of compute power and a huge amount of storage performance because the amount of data that needs to be manipulated and transferred to accomplish these clinical and efficiency outcomes is staggering," Rubens said.

Indeed, to successfully develop machine learning applications, developers must avoid "GPU starvation," which occurs when a processor falters because it can't access needed data. The gap between processing power and storage capacity, however, is widening. Consider the following: Between 2015 and 2017, the amount of compute power required by leading deep learning algorithms jumped 15 times. To meet this need, the compute

[1] Slabodkin, G. Experts tell Senate panel of AI"s growing potential. Health Data Management. https://www.healthdatamanagement.com/news/experts-tell-senate-panel-of-ais-growing-potential

[2] NVIDIA. Mayo Clinic Turns to AI to Improve Brain Tumor Treatment. https://blogs.nvidia.com/blog/2017/04/17/ai-to-predict-brain-tumor-genomics/

[3] Kubota, T. Algorithm better at diagnosing pneumonia than radiologists. https://m.phys.org/news/2017-11-algorithm-pneumonia-radiologists.html

[4] Bazzoli, F. Researchers to link images, phenomes to classify prostate cancer cases. Health Data Management. https://www.healthdatamanagement.com/news/researchers-to-link-images-phenomes-to-classify-prostate-cancer-cases

power delivered by GPUs (processes that deliver 10 to 100 times the performance of a CPU) increased by a factor of 10. Unfortunately, though, the legacy storage capacity stood still at zero growth.[5]

The result: GPUs are left starving for data – making it difficult if not prohibitive – for organizations to develop the machine learning applications that can have a significant impact on quality and efficiency in the healthcare industry.

"So, you have the computing power, but if the GPUs are starved for data because your storage can't receive data fast enough, the whole thing basically falls flat, especially if researchers have to run things over and over, which is required when developing machine learning applications in medical imaging. We're talking about hundreds of terabytes or petabytes every time they do something and it can take days to copy hundreds of terabytes," Rubens said.

## Meeting Infrastructure Needs with an Advanced Storage Solution

While traditional storage platforms fail to provide what's needed, Pure Storage's FlashBlade is capable of delivering high performance access to billions of objects and files for tens of thousands of clients in parallel. FlashBlade can be expanded to 75 blades and deliver linear-scaling performance up to:

- 75GB/sec read throughput
- 25GB/sec write throughput
- 7.5M IOPS at 8PB capacity (assuming 3:1 compression)

In addition, FlashBlade can keep nourishing AI and machine learning efforts by making it possible to efficiently train the machine learning model -- even with the hundreds of terabytes required with medical imaging.

"When researchers bring these machine learning techniques to clinical practice and healthcare, they need to be run a vast amount of data through those training models. If they have storage that is not capable of keeping up, this either becomes impossible or impractical," Rubens pointed out.

In fact, such storage performance is especially important with neural networks, a system of hardware and/or software patterned after the operation of neurons in the human brain. Neural networks are a variety of deep learning technologies that often focus on solving complex signal processing or pattern recognition problems.

"When you have neural networks, the most important aspect is training the network to label data. Labeling data is a huge skill. Of course, if you have a set of images of dogs and cats, anyone can label the images correctly," Rubens pointed out. "In medical imaging, you need MDs or PhDs, extremely specialized people to do labeling. They basically run these labeled data sets through the neural networks, but unless they can feed the data quickly enough through the neural networks, they just are not able to get the accuracy of the network up to where it needs to be."

The FlashBlade storage platform is also an evergreen solution. As such, users never need to perform a "forklift" replacement. Because Flashblade delivers storage infrastructure as a service, capacity continually grows with burgeoning needs.

"Normally with traditional computing storage, developers buy some kind of infrastructure and a few years later, they have to replace everything. Essentially, with FlashBlade, the storage devices or platforms get better over time and there is no need to replace it with an updated model," Rubens said. "Not only is there no depreciation but with FlashBlade, the engine can be replaced as users are 'driving down I-95.' The users of the storage don't notice because it's totally transparent."

The platform also makes it easy for end-users to reach needed performance levels. In fact, no fine-tuning is required as parameters and balance loads are continually adjusted. "We don't want people to have to read through thousands of pages of manuals just to get the best performance they can. There are no knobs. There are no dials. There's nothing. All the complexity is behind the scenes. we actually don't have manuals because there is no reason to do that," Rubens said.

In the final analysis, the FlashBlade platform provides the storage capacity that is on par with the compute power needed to support the development of machine learning medical imaging applications. As such, developers and healthcare organizations can move beyond simply exploring AI's potential and can actually bring advanced machine learning applications to the market, and into everyday clinical practice. As such, patients can experience the very real quality and efficiency improvements associated with machine learning.

[5] Kim, R. Why the AI Industry Needs to Rethink Storage. https://blog.purestorage.com/ai-industry-needs-rethink-storage/