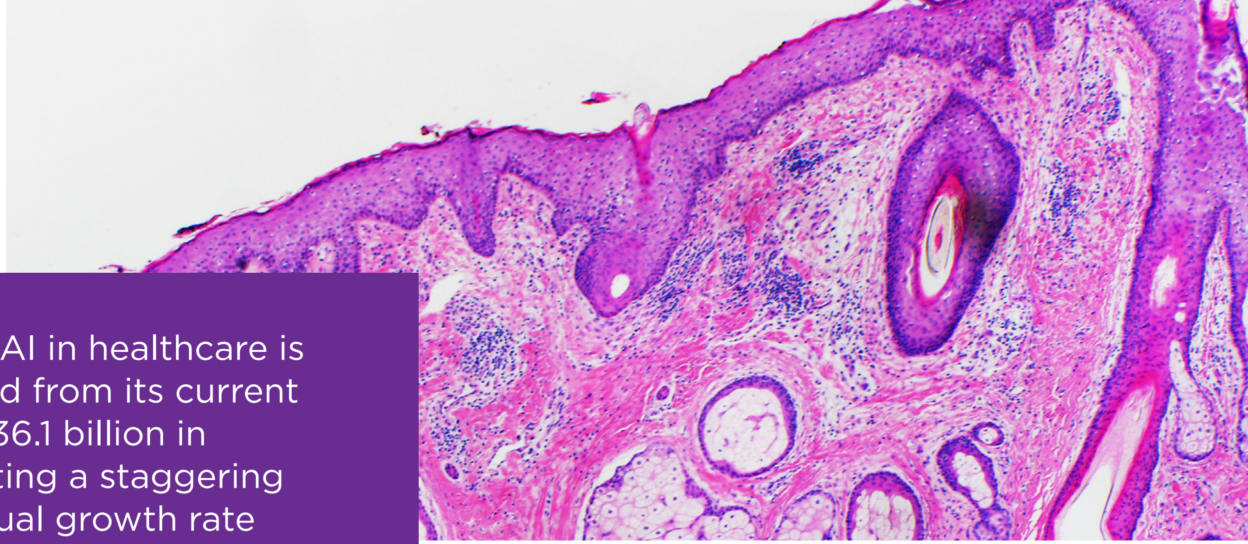
Computational Pathology Making Important Advancements

Convolutional neural networks proving effective

NetApp® | NVIDIA®

AI in Digital Pathology



The market for AI in healthcare is slated to expand from its current $2.1 billion to $36.1 billion in 2025, representing a staggering compound annual growth rate (CAGR) of 50.2%.[2]

Artificial intelligence has been slower to take off in the field of pathology than in other areas of medicine. But that might be about to change. Experts are starting to acknowledge that the many required pieces are in place, and the path forward for AI is bright.

"The future paradigm of pathology will be digital," writes an international team of researchers. "Instead of conventional microscopy, a pathologist will perform a diagnosis through interacting with images on computer screens and performing quantitative analysis."[1]

But other than in research settings, the road to fully digitized clinical pathology departments—incorporating whole slide imaging, convolutional neural networks, and cost-effective, high-performance computing and data storage—is still a work in progress. The reasons are both complex and pragmatic. The technology, for the most part, is still emerging. The best practices surrounding AI can vary. At the same time, data issues can create roadblocks.

Nonetheless, progress is being made. The digitization trend in pathology is undeniably accelerating. Productivity gains, as well as new insights into cancer detection and other abnormalities, are helping increase overall enthusiasm and acceptance of AI.

This paper reviews the current challenges but also discusses recent technical advances that can contribute to better solutions in digital pathology.

Approximately $18 billion of the AI market will be focused on deep learning technologies, according to estimates by Market Research Engine.

Background

In conventional pathology, slides are prepared from patient tissue samples and then reviewed by a pathologist under a microscope at high magnification. This manual process can be error prone, time consuming, and limiting when the pathologist needs to consult with an outside expert.

This is where digital pathology offers assistance. After tissue samples are prepared, the resulting slides are scanned with whole slide imaging (WSI) technology. A pathologist can then review and manipulate slide images on a high-resolution computer monitor rather than looking at the slides directly through a microscope. Pathologists can share images and collaborate with a few mouse clicks, easily perform critical measurements on screen, and compare patient images with reference images from very large underlying databases.

Pathology labs are also increasingly able to use computational pathology, applying various numerical and machine learning techniques to digital slide images to increase the speed, accuracy, and efficiency of diagnosis.

Challenge

The demand for pathology services is growing faster than the number of pathologists, resulting in a deepening shortage in many countries. With aging populations and increased cancer screening programs driving up workloads, pathology labs have to streamline operations to handle more cases in less time. Simultaneously, the rate of complex pathology exams is increasing, driving up the average time per case.

Meanwhile, pathology has been relatively slow to digitize. In radiology, digitization has offered a clear path to cost reduction and increased workflow efficiency over the past 3 decades by replacing printed film costs. However, digital pathology adds digital technology on top of existing physical processes, so cost benefits are less clear. Physical slides must still be prepared and stained, and they must be archived for extended periods. As a result, the prerequisites to computational pathology are seldom met, and fewer annotated digital pathology datasets exist for AI experimentation.

When physical slides are digitized, the size and complexity of the images pose an extra challenge. A single case might involve multiple images totaling from 0.5GB to 6GB, depending on the level of magnification. A single image might be several thousand pixels in length and width—and captured in full color. Rather than having deep learning algorithms operate on a whole image, it's common to divide the image into patches and analyze each patch. This approach adds significantly to the workload for both training and inference. These characteristics prevent the direct application of traditional deep learning approaches, mainly because very few deep learning use cases deal with such high-resolution images. Specific techniques such as intense preprocessing, tiling, and planning are required before these slides can be used for training computational pathology models.

Despite many deep learning–based algorithms now available for research and even clinical use, each trained for a specific use case, it can be difficult to determine how to meaningfully incorporate AI into a standard pathology workflow. The size of the data files makes it highly unlikely that labs will be able to use cloud resources for inference. Medium and large labs will need high-speed compute and data storage to support AI inference on the edge.

> "The future paradigm of pathology will be digital. Instead of conventional microscopy, a pathologist will perform a diagnosis through interacting with images on computer screens and performing quantitative analysis."
>
> —New Trends of Emerging Technologies in Digital Pathology[6]

Solutions and Benefits

Assuming organizations are ready to make the jump to digitize their pathology slides and workflows, there are many successful solutions that are improving outcomes. One is the capability for telepathology, which can help distribute and centralize workloads dynamically as needed. Computational pathology is another one, with many proven machine learning approaches that are improving the accuracy and automation of slide analysis. Convolutional neural networks (CNNs) are now considered the state-of-the-art way to build the best decision-making workflows in digital pathology.[3] Although pathologists are the only ones who can make a cancer diagnosis, CNNs help increase the accuracy and efficiency of a diagnosis and they help doctors identify benign or normal tissue more quickly, which can reduce unnecessary human intervention. For example, a research team in the Netherlands found that in prostate and breast cancer cases, 30% to 40% of the slides could be excluded automatically, freeing up valuable time for additional analysis.[4]

To support CNNs in practice, organizations need a variety of hardware, software, and infrastructure, and the advances have been rapid. It was not until 2017 that the FDA approved the first WSI system for primary diagnostics, so the growth has been remarkable.[5] Whereas tissue staining and slide preparation are unaffected by digitization, there has been an increase in the digitization of pathology and diagnostic labs, enabling pathologists to transform their workflow. They can integrate digital scanners with laboratory IT systems to better manage digital slides inside and outside an organization. With this workflow, they can review digital slides on screen rather than using a microscope, and they can report cases in an entirely digital workspace.

Advancements in the computational power and memory bandwidth of GPUs are continually reducing the compute-related bottlenecks of computational pathology.

Despite these advancements, the efficient implementation of WSI remains complex. Given the large file sizes, most deep learning methods don't use the whole image as input. Instead, they generate equal-sized patches from it in a preprocessing step and later feed these patches for training. Each high-resolution image can generate several thousands of patches, depending on the patch extraction approach.

Patch generation can be a bottleneck and can slow down the whole process. To mitigate this slowdown, organizations can process each image in parallel and accelerate this step of the WSI analysis pipeline. On a single server, multiple cores (multiprocessing) can be used to speed up the process further, but you're likely to reach the limits of the resources. To get the most out of existing and new data that's flowing in, you need to process data as quickly and efficiently as possible.

To do that, organizations can use distributed high-performance computing (HPC) frameworks, such as Apache Spark, to accelerate WSI preprocessing and to generate patches at scale by using multiple compute nodes. Multiple servers are processing data, so the question becomes: "What is the role of storage in such a high-data-demand case?" If data access needs are not met properly, storage can easily become a bottleneck, and compute nodes might starve for input data without being able to use resources to their maximum potential.

To support such high-performance I/O requirements, organizations can use BeeGFS, a parallel HPC file system. NetApp® E-Series storage with BeeGFS gives you consistent, near-real-time access to your data. To prevent bottlenecks and to support continuous high-performance workloads like AI, BeeGFS transparently spreads data across multiple servers and their back-end storage. And in addition to being open source (the basic features), BeeGFS comes with graphical administration and monitoring, unlike complex legacy open-source parallel file systems.

The fact that BeeGFS is clustered indicates that the server nodes work together to deliver a single file system that can be simultaneously mounted and accessed by other server nodes, commonly known as clients. The main takeaway is that clients can see and consume this distributed file system similarly to a local file system. Therefore, data can be accessed as is by native machine learning and deep learning frameworks. There's no need for extra tools like data movers, and you aren't required to write manual scripts.

To see how high-performance and low-latency E-Series storage facilitates WSI analysis with Apache Spark and BeeGFS, you can find the setup instructions and code used for this demonstration on GitHub.

Continued development of deep learning–based applications is likely to stimulate the growth of the medical imaging analytics market, expected to cross $2 billion by 2023, according to Signify Research.

Conclusion

The growth of AI in healthcare and pathology is undeniable, making pathology a recognized target for intensive AI development, particularly in the field of oncology and tissue biomarker analytics. The generation of high-resolution digital images—and the intricate, complex patterns required for disease recognition—provides unprecedented opportunity to apply AI in pathology for better patient outcomes.

To continue this work, organizations need to further refine the necessary training models and datasets, optimize workloads, and rely on evolving technical solutions that will bring costs down while improving speed and efficiency.

But the fact remains there is an ongoing convergence of innovative technologies that is transforming pathology, and patients are better for it.

About NetApp

NetApp is the leader in cloud data services, empowering global organizations to change their world with data. Together with our partners, we are the only ones who can help you build your unique data fabric. Simplify hybrid multicloud and securely deliver the right data, services, and applications to the right people at the right time. Learn more at **www.netapp.com**.

1. “New Trends of Emerging Technologies in Digital Pathology.” https://pubmed.ncbi.nlm.nih.gov/27100343/
2. “Artificial Intelligence in Healthcare Spending to Hit $36B.” https://healthitanalytics.com/news/artificial-intelligence-in-healthcare-spending-to-hit-36b
3. “Convolutional Neural Networks Can Accurately Distinguish Four Histologic Growth Patterns of Lung Adenocarcinoma in Digital Slides.” https://www.nature.com/articles/s41598-018-37638-9
4. “Deep Learning as a Tool for Increased Accuracy and Efficiency of Histopathological Diagnosis.” https://ui.adsabs.harvard.edu/abs/2016NatSR...626286L/abstract
5. “Translational AI and Deep Learning in Diagnostic Pathology.” https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6779702/pdf/fmed-06-00185.pdf
6. Gloria Bueno, M Milagro Fernández-Carrobles, Oscar Deniz, and Marcial García-Rojo, “New Trends of Emerging Technologies in Digital Pathology,” PubMed, https://pubmed. ncbi.nlm.nih.gov/27100343/.

NetApp®