

# Artificial intelligence decodes cancer pathology images without image gallery

Novartis researchers are collaborating with tech startup PathAI to search for hidden information in pathology slides.

By [Elizabeth Dougherty](#) | Nov 12, 2018

For 150 years, pathologists have been looking through microscopes at tissue samples mounted on slides to diagnose cancer. Each assessment is weighty: Does this patient have cancer or not?

The job of a pathologist is daunting. A single slide could contain hundreds of thousands of cells. Only a handful might be cancer. Inaccurate diagnosis rates range from 3-9% of cases, according to a recent [review](#).

Enter artificial intelligence (AI), an extra set of unbiased, indefatigable artificial eyes that could help catch errors. Many researchers are pursuing this possibility, but Novartis pathologists think AI might have an additional role to play. They hypothesize that pathology slides could contain information that helps explain why some patients respond to therapy when other seemingly similar patients do not.

To explore this idea, pathologists and data scientists from Novartis have joined forces with tech startup PathAI. They are training an AI system developed by PathAI to learn to see the same patterns pathologists see and then building on that to determine if the system can detect hidden but informative patterns too subtle or complex for pathologists to discern. The effort is part of a larger effort at Novartis to leverage data and digital technologies in ways that could help drug developers get the right drugs to the right patients faster.

In a first phase of testing, the collaborative team has trained the PathAI system to look at slides from untreated patients and distinguish tumor from normal tissue. The system can also identify different cell types on a slide reliably. For a pathologist, these feats are akin to finding a needle in a haystack and then labeling every piece of straw.

The ability to label every cell is becoming increasingly important as cancer therapies evolve to include medicines that target not only cancer cells but also immune cells. If computers can analyze an entire slide at once and quantify cell types and locations, they could potentially reveal patterns that predict how well a patient might fare on a given therapy.

“Hopefully we can figure out which features correlate with survival or response to a drug,” says Meg McLaughlin, a pathologist and Director of the Oncology Pathology and Biomarkers group in the Oncology Translational Research team at the Novartis Institutes for BioMedical Research (NIBR).

With a recent explosion of experimental immuno-oncology options alongside therapies that target cancer-driving mutations, one of the biggest challenges for drug hunters is matching the most appropriate therapy to individual patients. While genomic information helps drive smart decisions, valuable clues in pathology slides could also help. “We want to create a platform that enables the field of pathology to support the accelerating pace of drug development,” says Andrew Beck, a pathologist, computer scientist and CEO of PathAI, located in Boston, Massachusetts, in the US.

We want to create a platform that enables the field of pathology to support the accelerating pace of

drug development.

**Andrew Beck**, CEO of PathAI

## Training the AI model

In collaboration with the Institute of Pathology at the University Hospital Basel in Switzerland, the Novartis team gained access to 400 pathology images from breast and lung cancer tissues along with anonymized information about the patients' diagnoses and survival times.

The challenge for PathAI's platform? Given an image, identify cancer, identify cell types and predict the patient's probability of surviving five years.

One way to approach the challenge is to feed a set of untrained AI algorithms a subset of the data and see what it learns. Unlike a trained pathologist, the machine approaches the problem with no knowledge of cells or cancer.

"A human already has a lot of knowledge," says NIBR data scientist Holger Hoefling, who is working on the project with PathAI and with an internal NIBR group aiming to use AI to assess safety concerns in pathology images. "Think about autonomous cars. To train a car to drive, the amount of time and data required for training is gigantic. In contrast, you put a human behind the wheel for 20 hours and let them drive."

To give the untrained algorithms more knowledge about the training data, PathAI decided to feed them even more rich data. A team of consulting pathologists marks up the slides, giving the algorithms more information to work with. It's a bit like annotations in a hefty piece of literature that highlight and explain critical passages.

For example, when training the algorithms to distinguish cell types, PathAI diced the training slides into about 10 000 smaller images and had pathologists label the cell types in each slice. "We had to think really hard about how we annotate the images," says McLaughlin. "That step determines to a large extent what you get out of the AI model in the end."

## What is a black box?

AI experts refer to the trained algorithms as a "black box" because it's difficult to know what the system has learned from the training data or how it makes decisions.

Inside the black box is a set of machine learning algorithms. These algorithms are a cascade of formulas that recognize features, such as the presence of a certain shape, and associate them with real-world data, such as how long a patient actually survived.

As the algorithms see more and more images, they adjust their understanding of the patterns they see in the data. Eventually they learn that certain shapes in a slide predict likely health outcomes, such as having a good chance of living one year or a poor chance of surviving six months.

The black box approach has the benefit of taking a fresh view of the data, so it can reveal unexpected biological patterns. But it can also discover patterns that have no biological meaning at all. Data scientists need to scrutinize the AI model's output, identify the meaningless conclusions, and adjust the training data and algorithms in ways that weed them out.

## Seeing through a machine's eyes

After training, the PathAI platform lets users see pathology images through the machine's eyes. Regions of the slides determined to be cancer glow bright red in a field of green surrounding tissue. Different cell types stand out in vivid colors like candies in a dish. The existing platform is for research use only, but PathAI aims to build applications that could be used by doctors in the future.

Now that the researchers have shown that the PathAI system has the potential to see what pathologists see, they want to find out if there's information in those images that isn't obvious to pathologists.

For example, they wonder if the distribution and abundance of certain cells, such as immune cells, could hold clues about how well a patient might do on immune therapy. To find out using the human eye would require the painstaking scrutiny of tens of thousands of cells per slide, an implausible task. "I could potentially do it," says McLaughlin. "But it would take forever."

With AI, however, the task becomes feasible. McLaughlin and her team at Novartis are supplying PathAI with pathology images and data about survival times and response to therapy from a recent Novartis clinical trial for cancer. Gathering this data together and sharing it with PathAI is no small task. In addition to locating slides and verifying the consent of patients, the team must present orderly and consistent data from doctor's visits, which can be a challenge when multiple doctors working across several clinics collected it.

"The era we're heading into is more about data than it is about algorithms," says Lee Cooper, assistant professor of biomedical informatics and biomedical engineering at Emory University in the US. Cooper specializes in using machine learning to understand pathology images and is collaborating with NIBR researchers. "The algorithms will continue to improve, but really the challenge we're facing is how to produce the datasets we need to build the best algorithms we can."

Once the data is in hand, PathAI will have the images annotated. After that, it will be up to the machines to find any hidden messages.

"If we can show that this method can take in this data and overlay information we haven't seen before from pathologists, we'll be onto something of potential value," says Beck.

*Video by PJ Kaszas.*

## **Reimagine medicine with Novartis**

Find career opportunities in research at Novartis.

[Learn More](#)

---

**Source URL:** <https://prod1.novartis.com/stories/artificial-intelligence-decodes-cancer-pathology-images-without-image-gallery>

### **List of links present in page**

1. <https://prod1.novartis.com/stories/artificial-intelligence-decodes-cancer-pathology-images-without-image-gallery>
2. <https://prod1.novartis.com/tags/category/discovery>
3. <https://prod1.novartis.com/tags/authors/elizabeth-dougherty>
4. <https://jcp.bmj.com/content/early/2018/08/01/jclinpath-2018-205226>