

# Researchers develop AI-driven tool for near real-time cancer surveillance

March 13 2024, by Mark Alewine

---



Cancer cell during cell division. Credit: National Institutes of Health

Artificial intelligence has delivered a major win for pathologists and researchers in the fight for improved cancer treatments and diagnoses.

In partnership with the National Cancer Institute, or NCI, researchers from the Department of Energy's Oak Ridge National Laboratory and Louisiana State University developed a long-sequenced AI transformer capable of processing millions of pathology reports to provide experts researching [cancer](#) diagnoses and management with exponentially more accurate information on cancer reporting.

"Our goal is trying to see if we can automate the process of extraction of specific cancer site information from these pathology reports and make it into structured data for nation level cancer incidence reporting," said Mayanka Chandra Shekar, a research scientist in the Computational Sciences and Engineering Division at ORNL.

The team's work is [published](#) in *Clinical Cancer Informatics*.

AI transformer models are trained on large amounts of data and "transform" them into information that is useful and digestible to scientists. Using the secure CITADEL framework on the Oak Ridge Leadership Computing Summit supercomputer, researchers at ORNL used the specialized transformer model to process 2.7 million cancer pathology reports. This model, known as Path-BigBird, pulls data from six Surveillance, Epidemiology and End Results, or SEER cancer registries.

The NCI's SEER program is an authoritative source of information on cancer incidence and survival in the United States. SEER currently collects and publishes cancer incidence and survival data from population-based cancer registries covering approximately 48% of the U.S. population.

"We wanted to build a language model where we could ask, 'Can we build something that will understand the language of pathology and help us to create predictive modeling or information extraction models which

will basically extract cancer site, subsite and other key details out of pathology reports?" Chandrashekar said.

Currently, these cancer registries are updated by hand, leaving a two-year gap between the cancer incidence and its reporting, which means if there is an increase in cancer rate nationally, researchers have to wait two years before recognizing this area of concern.

By effectively processing the information from millions of pathology reports, Path-BigBird has the potential to streamline the speed and accuracy for pathology information extraction and outperform traditional deep learning approaches to gathering important information such as identifying cancer sites, histology and improve the precision of cancer incidence reporting at a population level.

"Our current deployed deep learning model has autocoded around 23% of reports processed by the cancer registries, saving researchers valuable time in their quest for near-real time cancer reporting," said Chandrashekar. She added that this advancement opens the door to creating a comprehensive model pathology language that can successfully perform tasks more rapidly than ever.

"Usage of this model opens up a whole new world," said Chandrashekar. "We can extend to extract biomarkers and other recurrent cancer issues using the same model because now it's able to understand pathology specific language. We can expand it beyond the focus of what we started," she added.

The turning point in the research came when the team realized a broader scope of language was needed for the AI model to operate more accurately. By incorporating more clinical language along with [pathology](#) reports, Chandrashekar and her team saw a dramatic improvement in both accuracy and performance.

"This gave us a scope to understand that having a limited vocabulary might limit us in understanding the nuances of the behavior in certain tasks," Chandrashekar said. "Meanwhile, including more vocabulary is going to create a better model to perform normal tasks, as well as the harder ones."

The inclusive language incorporated into the AI model was reflective of the wide range of researchers assembled for the team, who spent two years working on this project.

Chandrashekar added, "Our team included people from natural language processing experts, high-performance computing scientists and epidemiologists, so we were a group of completely interdisciplinary parts where we had to understand, 'What is being asked and can we run it securely at scale?'"

Researchers have tested the Path-BigBird model for essential information extraction tasks. Knowing the potential for transformer models from popular models such as BERT and GPT, they expect to extend and adapt for downstream tasks that are useful in population health, like entity recognition, the location of essential text and question-answering systems. The Path-BigBird model could also be a turning point by providing a clearer understanding of cancer trends and facilitating public health interventions for at-risk communities.

Chandrashekar said the team's attention has now moved on to implementing new tasks for the model to complete, such as identifying biomarkers, cancer recurrence rates and other aspects of cancer incidence reporting.

"We are trying to see if we can use a similar model which doesn't have to go through a lot of training and see how we can expand it across these things," she said. "And given the rate at which large [language](#) models are

being built by industry, we are trying to understand how we can leverage some of this knowledge to see if we can use existing models for our particular use case."

**More information:** Mayanka Chandrashekar et al, Path-BigBird: An AI-Driven Transformer Approach to Classification of Cancer Pathology Reports, *JCO Clinical Cancer Informatics* (2024). [DOI: 10.1200/CCI.23.00148](https://doi.org/10.1200/CCI.23.00148)

Provided by Oak Ridge National Laboratory

Citation: Researchers develop AI-driven tool for near real-time cancer surveillance (2024, March 13) retrieved 29 April 2024 from <https://medicalxpress.com/news/2024-03-ai-driven-tool-real-cancer.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.