Researchers develop an AI model that can detect future lung cancer risk

January 20 2023, by Alex Ouyang

Researchers from Massachusetts General Hospital and MIT stand in front of a CT scanner at MGH, where some of the validation data was generated. Left to right: Regina Barzilay, Lecia Sequist, Florian Fintelmann, Ignacio Fuentes, Peter Mikhael, Stefan Ringer, and Jeremy Wohlwend Credit: Guy Zylberberg.

The name Sybil has its origins in the oracles of Ancient Greece, also

known as sibyls: feminine figures who were relied upon to relay divine knowledge of the unseen and the omnipotent past, present, and future. Now, the name has been excavated from antiquity and bestowed on an artificial intelligence tool for lung cancer risk assessment being developed by researchers at MIT's Abdul Latif Jameel Clinic for Machine Learning in Health, Mass General Cancer Center (MGCC), and Chang Gung Memorial Hospital (CGMH).

Lung cancer is the No. 1 deadliest cancer in the world, resulting in 1.7 million deaths worldwide in 2020, killing more people than the next three deadliest cancers combined.

"It's the biggest cancer killer because it's relatively common and relatively hard to treat, especially once it has reached an advanced stage ," says Florian Fintelmann, MGCC thoracic interventional radiologist and coauthor on the new work. "In this case, it's important to know that if you detect lung cancer early, the long-term outcome is significantly better. Your five-year survival rate is closer to 70 percent, whereas if you detect it when it's advanced, the five-year survival rate is just short of 10 percent."

Although there has been a surge in new therapies introduced to combat lung cancer in recent years, the majority of patients with lung cancer still succumb to the disease. Low-dose computed tomography (LDCT) scans of the lung are currently the most common way patients are screened for lung cancer with the hope of finding it in the earliest stages, when it can still be surgically removed. Sybil takes the screening a step further, analyzing the LDCT image data without the assistance of a radiologist to predict the risk of a patient developing a future lung cancer within six years.

In their new paper published in the *Journal of Clinical Oncology*, Jameel Clinic, MGCC, and CGMH researchers demonstrated that Sybil

obtained C-indices of 0.75, 0.81, and 0.80 over the course of six years from diverse sets of lung LDCT scans taken from the National Lung Cancer Screening Trial (NLST), Mass General Hospital (MGH), and CGMH, respectively—models achieving a C-index score over 0.7 are considered good and over 0.8 is considered strong. The ROC-AUCs for one-year prediction using Sybil scored even higher, ranging from 0.86 to 0.94, with 1.00 being the highest score possible.

Despite its success, the 3D nature of lung CT scans made Sybil a challenge to build. Co-first-author Peter Mikhael, an MIT Ph.D. student in electrical engineering and computer science, and affiliate of Jameel Clinic and the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), likened the process to "trying to find a needle in a haystack."

The imaging data used to train Sybil was largely absent of any signs of cancer because early-stage lung cancer occupies small portions of the lung—just a fraction of the hundreds of thousands of pixels making up each CT scan. Denser portions of lung tissue are known as lung nodules, and while they have the potential to be cancerous, most are not, and can occur from healed infections or airborne irritants.

To ensure that Sybil would be able to accurately assess cancer risk, Fintelmann and his team labeled hundreds of CT scans with visible cancerous tumors that would be used to train Sybil before testing the model on CT scans without discernible signs of cancer.

MIT electrical engineering and computer science Ph.D. student Jeremy Wohlwend, co-author of the paper and Jameel Clinic and CSAIL affiliate, was surprised by how highly Sybil scored despite the lack of any visible cancer. "We found that while we [as humans] couldn't quite see where the cancer was, the model could still have some predictive power as to which lung would eventually develop cancer," he recalls.

"Knowing [Sybil] was able to highlight which side was the most likely side was really interesting to us."

Co-author Lecia V. Sequist, a medical oncologist, lung cancer expert, and director of the Center for Innovation in Early Cancer Detection at MGH, says the results the team achieved with Sybil are important "because lung cancer screening is not being deployed to its fullest potential in the U.S. or globally, and Sybil may be able to help us bridge this gap."

Lung cancer screening programs are underdeveloped in regions of the United States hardest hit by lung cancer due to a variety of factors. These range from stigma against smokers to political and policy landscape factors like Medicaid expansion, which varies from state to state.

Moreover, many patients diagnosed with lung cancer today have either never smoked or are former smokers who quit over 15 ago—traits that make both groups ineligible for lung cancer CT screening in the United States.

"Our training data consisted only of smokers because this was a necessary criterion for enrolling in the NLST," Mikhael says. "In Taiwan, they screen nonsmokers, so our validation data is expected to contain people who didn't smoke, and it was exciting to see Sybil generalize well to that population."

"An exciting next step in the research will be testing Sybil prospectively on people at risk for lung cancer who have not smoked or who quit decades ago," says Sequist. "I treat such patients every day in my lung cancer clinic and it's understandably hard for them to reconcile that they would not have been candidates to undergo screening. Perhaps that will change in the future."

There is a growing population of patients with lung cancer who are categorized as nonsmokers. Women nonsmokers are more likely to be diagnosed with lung cancer than men who are nonsmokers. Globally, over 50 percent of women diagnosed with lung cancer are nonsmokers, compared to 15 to 20 percent of men.

MIT Professor Regina Barzilay, a paper co-author and the Jameel Clinic AI faculty lead, who is also a member of the Koch Institute for Integrative Cancer Research, credits MIT and MGH's joint efforts on Sybil to Sylvia, the sister to a close friend of Barzilay and one of Sequist's patients.

"Sylvia was young, healthy and athletic—she never smoked," Barzilay recalls. "When she started coughing, neither her doctors nor her family initially suspected that the cause could be lung cancer. When Sylvia was finally diagnosed and met Dr. Sequist, the disease was too advanced to revert its course. When mourning Sylvia's death, we couldn't stop thinking how many other patients have similar trajectories."