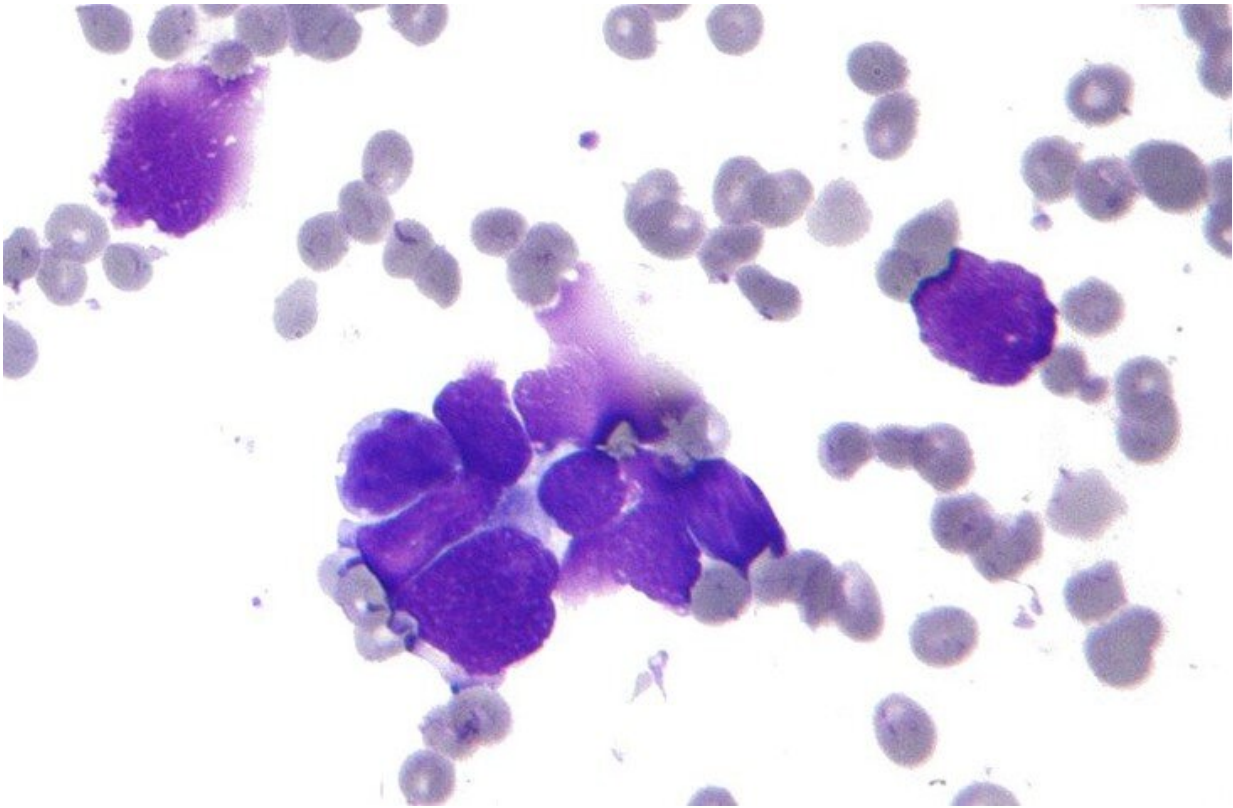# Deep learning may help doctors choose better lung cancer treatments

February 17 2021, by Matt Swayne

Deep learning, a powerful machine learning model, could guide doctors and healthcare workers in weighing treatment and care options, according to a team of Great Valley researchers. Credit: Wikimedia

Doctors and healthcare workers may one day use a machine learning model, called deep learning, to guide their treatment decisions for lung

cancer patients, according to a team of Penn State Great Valley researchers.

In a study, the researchers report that they developed a [deep learning](#) model that, in certain conditions, was more than 71% accurate in predicting survival expectancy of [lung cancer patients](#), significantly better than traditional machine learning models that the team tested. The other machine learning models the team tested had about a 61% accuracy rate.

Information on a patient's survival expectancy could help guide doctors and caregivers in making better decisions on using medicines, allocating resources and determining the intensity of care for patients, according to Youakim Badr, associate professor of data analytics.

"This is a high-performance system that is highly accurate and is aimed at helping doctors make these important decisions about providing care to their patients," said Badr. "Of course, this tool can't be used as a substitute for a doctor in making decisions on lung [cancer](#) treatments."

According to Robin G. Qiu, professor of information science and engineering and an affiliate of the Institute for Computational and Data Sciences, the model can analyze a large amount of data—typically called features in machine learning—that describe the patients and the disease to understand how a combination of factors affect lung cancer survival periods. Features can include information such as types of cancer, size of tumors, the speed of tumor growth, and demographic data.

What is deep learning?

Deep learning may be uniquely suited to tackle lung cancer prognosis because the model can provide the robust analysis necessary in cancer research, according to the researchers, who report their findings in

International Journal of Medical Informatics. Deep learning is a type of machine learning that is based on artificial neural networks, which are generally modeled on how the human brain's own neural network functions.

In deep learning, however, developers apply a sophisticated structure of multiple layers of these artificial neurons, which is why the model is referred to as "deep." The learning aspect of deep learning comes from how the system learns from connections between data and labels, said Badr.

"Deep learning is a machine-learning algorithm that makes associations between the data, itself, and the labels that we use to describe the data examples," said Badr. "By making these associations, it learns from the data."

Qiu added that deep learning's structure offers several advantages for many data science tasks, especially when confronted with data sets that have a large number of records—in this case, patients—as well as a large number of features.

"It improves performance tremendously," said Qiu. "In deep learning we can go deeper, which is why they call it that. In traditional machine learning, you have a simple structure of layers of neural networks. In each layer, you have a group of cells. In deep learning, there are many layers of these cells that can be architected into a sophisticated structure to perform better feature transformation and extraction, which gives you the ability to further improve the accuracy of any model."

In the future, the researchers would like to improve the model and test its ability to analyze other types of cancers and medical conditions.

"The accuracy rate is good, but it's not perfect, so part of our future

work is to improve the model," said Qiu.

To further improve their [deep learning model](#), the researchers would also need to connect with domain experts, who are people who have specific knowledge. In this case, the researchers would like to connect with experts on specific cancers and medical conditions.

"In a lot of cases, we might not know a lot of features that should go into the model," said Qiu. "But, by collaborating with domain experts, they could help us collect important features about patients that we might not be aware of and that would further improve the [model](#)."

The researchers analyzed data from the Surveillance, Epidemiology, and End Results (SEER) program. The SEER dataset is one of the biggest and most comprehensive databases on the early diagnosis information for cancer patients in the United States, according to Shreyesh Doppalapudi, a graduate-student research assistant and first author of the paper. The program's cancer registries cover almost 35% of U.S. cancer patients.

"One of the really good things about this data is that it covers a large section of the population and it's really diverse," said Doppalapudi. "Another good thing is that it covers a lot of different features, which you can use for many different purposes. This becomes very valuable, especially when using machine learning approaches."

Doppalapudi added that the team compared several deep learning approaches, including [artificial neural networks](#), convolutional neural networks and recurrent neural networks, to traditional machine learning models. The deep learning approaches performed much better than the traditional machine learning methods, he said.

Deep learning architecture is better suited to processing such large,

diverse datasets, such as the SEER program, according to Doppalapudi. Working on these types of datasets requires robust computational capacity. In this study, the researchers relied on ICDS's Roar supercomputer.

With about 800,000 to 900,000 entries in the SEER dataset, the researchers said that manually finding these associations in the data with an entire team of medical researchers would be extremely difficult without assistance from machine learning.

If it were only three fields I would say it would be impossible—and we had about 150 fields," said Doppalapudi. "Understanding all of those different fields and then reading and learning from that information, would be impossible."

  More information: Shreyesh Doppalapudi et al. Lung cancer survival period prediction and understanding: Deep learning approaches, *International Journal of Medical Informatics* (2020). DOI: 10.1016/j.ijmedinf.2020.104371

Provided by Pennsylvania State University