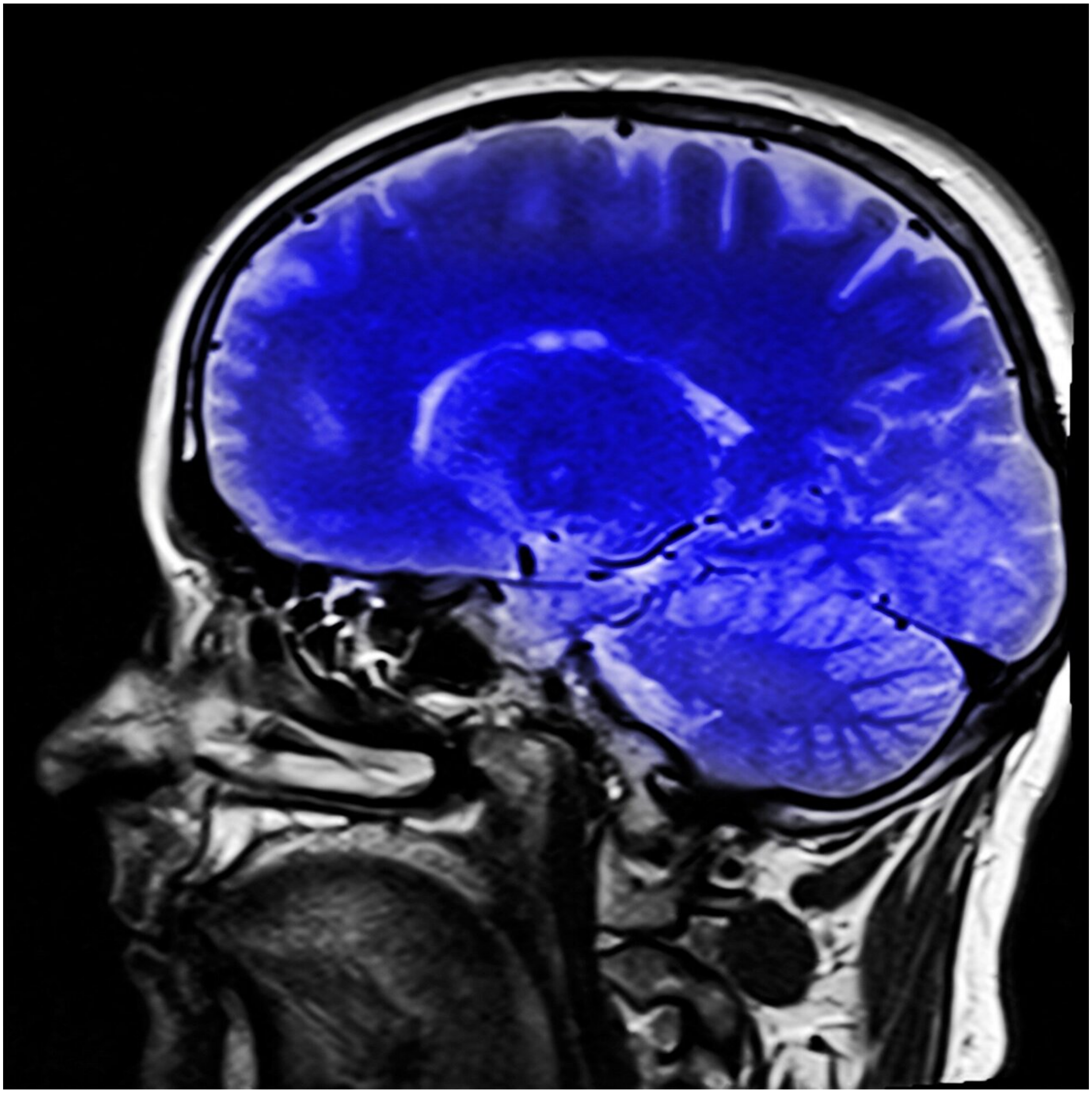
Medical Xpress

AI models to analyze cancer images take shortcuts that introduce bias

July 22 2021



Credit: CC0 Public Domain

Artificial intelligence tools and deep learning models are a powerful tool in cancer treatment. They can be used to analyze digital images of tumor biopsy samples, helping physicians quickly classify the type of cancer, predict prognosis and guide a course of treatment for the patient. However, unless these algorithms are properly calibrated, they can sometimes make inaccurate or biased predictions.

A new study led by researchers from the University of Chicago shows that deep learning models trained on large sets of cancer genetic and tissue histology data can easily identify the institution that submitted the images. The models, which use machine learning methods to "teach" themselves how to recognize certain cancer signatures, end up using the submitting site as a shortcut to predicting outcomes for the patient, lumping them together with other patients from the same location instead of relying on the biology of individual patients. This in turn may lead to bias and missed opportunities for treatment in patients from racial or ethnic minority groups who may be more likely to be represented in certain medical centers and already struggle with access to care.

"We identified a glaring hole in the in the current methodology for deep learning model development which makes certain regions and patient populations more susceptible to be included in inaccurate algorithmic predictions," said Alexander Pearson, MD, Ph.D., assistant Assistant Professor of Medicine at UChicago Medicine and co-senior author. The study was published July 20, in *Nature Communications*.

One of the first steps in treatment for a cancer patient is taking a biopsy, or small tissue sample of a tumor. A very thin slice of the tumor is affixed to glass slide, which is stained with multicolored dyes for review by a pathologist to make a diagnosis. Digital images can then be created for storage and remote analysis by using a scanning microscope. While these steps are mostly standard across pathology labs, minor variations in the color or amount of stain, tissue processing techniques and in the imaging equipment can create unique signatures, like tags, on each image. These location-specific signatures aren't visible to the naked eye, but are easily detected by powerful deep learning algorithms.

These algorithms have the potential to be a valuable tool for allowing physicians to quickly analyze a tumor and guide treatment options, but the introduction of this kind of bias means that the models aren't always basing their analysis on the biological signatures it sees in the images, but rather the image artifacts generated by differences between submitting sites.

Pearson and his colleagues studied the performance of deep learning models trained on data from the Cancer Genome Atlas, one of the largest repositories of cancer genetic and tissue image data. These models can predict survival rates, gene expression patterns, mutations, and more from the tissue histology, but the frequency of these patient characteristics varies widely depending on which institutions submitted the images, and the model often defaults to the "easiest" way to distinguish between samples—in this case, the submitting site.

For example, if Hospital A serves mostly affluent patients with more resources and better access to care, the images submitted from that hospital will generally indicate better patient outcomes and survival rates. If Hospital B serves a more disadvantaged population that struggles with access to quality care, the images that site submitted will generally predict worse outcomes.

The research team found that once the models identified which institution submitted the images, they tended to use that as a stand in for other characteristics of the image, including ancestry. In other words, if the staining or imaging techniques for a slide looked like it was submitted by Hospital A, the models would predict better outcomes, whereas they would predict worse outcomes if it looked like an image from Hospital B. Conversely, if all patients in Hospital B had biological characteristics based on genetics that indicated a worse prognosis, the algorithm would link the worse outcomes to Hospital B's staining patterns instead of things it saw in the tissue.

"Algorithms are designed to find a signal to differentiate between images, and it does so lazily by identifying the site," Pearson said. "We actually want to understand what biology within a tumor is more likely to predispose resistance to treatment or early metastatic disease, so we have to disentangle that site-specific digital histology signature from the true biological signal."

The key to avoiding this kind of bias is to carefully consider the data used to train the models. Developers can make sure that different disease outcomes are distributed evenly across all sites used in the training data, or by isolating a certain site while training or testing the model when the distribution of outcomes is unequal. The result will produce more accurate tools that can get physicians the information they need to quickly diagnose and plan treatments for cancer patients.

"The promise of artificial intelligence is the ability to bring accurate and rapid precision health to more people," Pearson said. "In order to meet the needs of the disenfranchised members of our society, however, we have to be able to develop algorithms which are competent and make relevant predictions for everyone."

More information: Frederick M. Howard et al, The impact of site-specific digital histology signatures on deep learning model accuracy and bias, *Nature Communications* (2021). DOI: 10.1038/s41467-021-24698-1

Provided by University of Chicago Medical Center

Citation: AI models to analyze cancer images take shortcuts that introduce bias (2021, July 22) retrieved 24 April 2024 from https://medicalxpress.com/news/2021-07-ai-cancer-images-shortcuts-bias.html