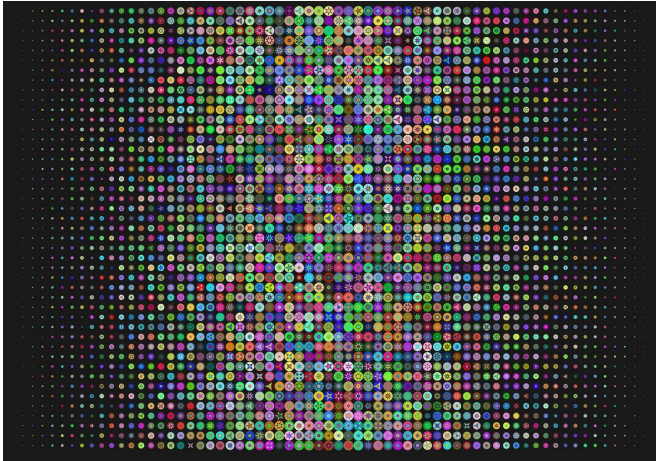# New machine learning method allows hospitals to share patient data—privately

28 July 2020

Credit: CC0 Public Domain

To answer medical questions that can be applied to a wide patient population, machine learning models rely on large, diverse datasets from a variety of institutions. However, health systems and hospitals are often resistant to sharing patient data, due to legal, privacy, and cultural challenges.

An emerging technique called federated learning is a solution to this dilemma, according to a study published Tuesday in the journal *Scientific Reports*, led by senior author Spyridon Bakas, Ph.D., an instructor of Radiology and Pathology & Laboratory Medicine in the Perelman School of Medicine at the University of Pennsylvania.

Federated learning—an approach first implemented by Google for keyboards' autocorrect functionality—trains an algorithm across multiple decentralized devices or servers holding local data samples, without exchanging them. While the approach could potentially be used to answer many different medical questions, Penn Medicine researchers have shown that federated learning is

successful specifically in the context of brain imaging, by being able to analyze magnetic resonance imaging (MRI) scans of brain tumor patients and distinguish healthy brain tissue from cancerous regions.

A model trained at Penn Medicine, for example, can be distributed to hospitals around the world. Doctors can then train on top of this shared model, by inputting their own patient brain scans. Their new model will then be transferred to a centralized server. The models will eventually be reconciled into a consensus model that has gained knowledge from each of the hospitals, and is therefore clinically useful.

"The more data the computational model sees, the better it learns the problem, and the better it can address the question that it was designed to answer," Bakas said. "Traditionally, machine learning has used data from a single institution, and then it became apparent that those models do not perform or generalize well on data from other institutions."

The federated learning model will need to be validated and approved by the U.S. Food and Drug Administration before it can be licensed and commercialized as a clinical tool for physicians. But if and when the model is commercialized, it would help radiologists, radiation oncologists, and neurosurgeons make important decisions about patient care, Bakas said. Nearly 80,000 people will be diagnosed with a brain tumor this year, according to the American Brain Tumor Association.

"Studies have shown that, when it comes to tumor boundaries, not only can different physicians have different opinions, but the same physician assessing the same scan can see different tumor boundary definition on one day of the week versus the next," he said. "Artificial Intelligence allows a physician to have more precise information about

where a tumor ends, which directly affects a patient's treatment and prognosis."

To test the effectiveness of federated learning and compare it to other machine learning methods, Bakas collaborated with researchers from the University of Texas MD Anderson Cancer Center, Washington University, and the Hillman Cancer Center at the University of Pittsburgh, while Intel Corporation contributed privacy-protecting software to the project.

The study began with a model that was pre-trained on multi-institutional data from an open-source repository known as the International Brain Tumor Segmentation, or BraTS, challenge. BraTS currently provides a dataset that includes more than 2,600 brain scans captured with magnetic resonance imaging (MRI) from 660 patients. Next, 10 hospitals participated in the study by training AI models with their own patient data. The federated learning technique was then used to aggregate the data and create the consensus model.

The researchers compared federated learning to models trained by single institutions, as well as to other collaborative-learning approaches. The effectiveness of each method was measured by testing them against scans that were annotated manually by neurologists. When compared to a model trained with centralized data that did not protect patient privacy, federated learning was able to perform almost (99 percent) identically. The findings also indicated that increased access to data through data private, multi-institutional collaborations can benefit model performance.

The findings from this study have paved the way for a much larger, ambitious collaboration between Penn Medicine, Intel, and 30 partner institutions, supported by a $1.2 million grant from the National Cancer Institute of the National Institutes of Health that was awarded to Bakas earlier this year. Intel announced in May that Bakas will lead the project, in which the 30 institutions, across nine countries, will use the federated learning approach to train a consensus AI model on brain tumor data. The final goal of the project will be to create an open-source tool for any clinician at any hospital to use. The development of the tool in Penn's Center for

Biomedical Image Computing & Analytics (CBICA) is being led by senior software developer Sarthak Pati, MS.

Study co-author Rivka Colen, MD, an associate professor of Radiology at the University of Pittsburgh School of Medicine, said that this paper and the larger federated learning project open up possibilities for even more uses of Artificial Intelligence in health care.

"I think it's a huge game changer," Colen said. "Radiomics is to radiology what genomics was to pathology. AI will revolutionize this field, because, right now, as a radiologist, most of what we do is descriptive. With deep learning, we're able to extract information that is hidden in this layer of digitized images."