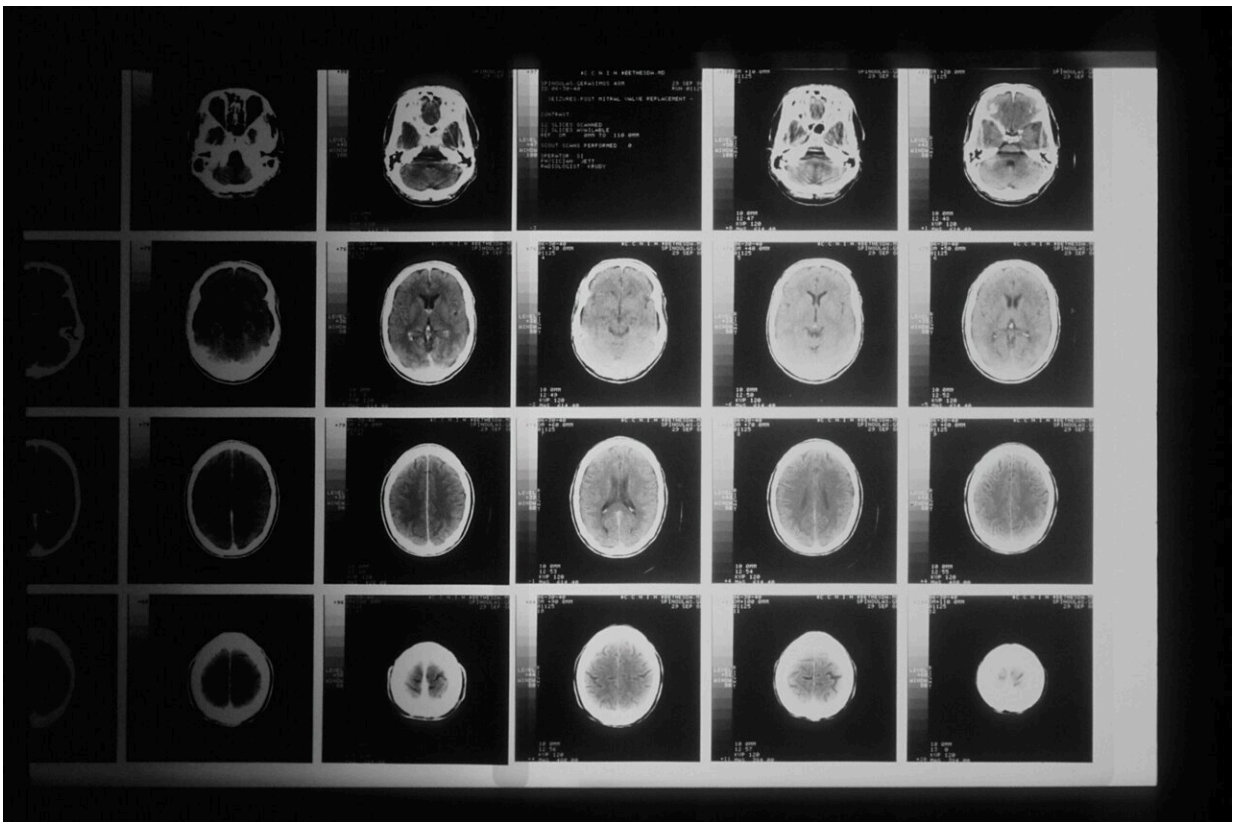


Federated machine learning enables the largest brain tumor study to-date, without sharing patient data

December 5 2022



Credit: Unsplash/CC0 Public Domain

Researchers at Penn Medicine and Intel Corporation led the largest-to-date global machine learning effort to securely aggregate knowledge

from brain scans of 6,314 glioblastoma (GBM) patients at 71 sites around the globe and develop a model that can enhance identification and prediction of boundaries in three tumor sub-compartments, without compromising patient privacy. Their findings were published today in *Nature Communications*.

"This is the single largest and most diverse dataset of glioblastoma patients ever considered in the literature, and was made possible through federated learning," said senior author Spyridon Bakas, Ph.D., an assistant professor of Pathology & Laboratory Medicine, and Radiology, at the Perelman School of Medicine at the University of Pennsylvania. "The more data we can feed into machine learning models, the more accurate they become, which in turn can improve our ability to understand, treat, and remove glioblastoma in patients with more precision."

Researchers studying rare conditions, like GBM, an aggressive type of brain tumor, often have patient populations limited to their own institution or geographical location. Due to privacy protection legislation, such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA) in the United States, and General Data Protection Regulation (GDPR) in Europe, data sharing collaborations across institutions without compromising [patient privacy](#) data is a major obstacle for many healthcare providers.

A newer machine learning approach, called federated learning, offers a solution to these hurdles by bringing the machine learning algorithm to the data instead of following the current paradigm of centralizing data to the algorithms. Federated learning—an approach first implemented by Google for keyboards' autocorrect functionality—trains a machine learning algorithm across multiple decentralized devices or servers (in this case, institutions) holding local data samples, without actually exchanging them. It has been [previously shown](#) to allow clinicians at

institutions in different countries to collaborate on research without sharing any private patient data.

Bakas led this massive collaborative study along with first authors Sarthak Pati, MS, a senior software developer at Penn's Center for Biomedical Image Computing & Analytics (CBICA), Ujjwal Baid, Ph.D., a postdoctoral researcher at CBICA, Brandon Edwards, Ph.D., a research scientist at Intel Labs, and Micah Sheller, a research scientist at Intel Labs.

"Data helps to drive discovery, especially in rare cancers where available data can be scarce. The federated approach we outline allows for access to maximal data while lowering institutional burdens to data sharing." said Jill Barnholtz-Sloan, Ph.D., adjunct Professor at Case Western Reserve University School of Medicine.

The [model](#) followed a staged approach. The first stage, called a *public initial model*, was pre-trained using publicly available data from the [International Brain Tumor Segmentation \(BraTS\) challenge](#). The model was tasked with identifying boundaries of three GBM tumor sub-compartments: "enhancing tumor" (ET), representing the vascular blood-brain barrier breakdown within the tumor; the "tumor core" (TC), which includes the ET and the part which kills tissue, and represents the part of the tumor relevant for surgeons who remove them; and the "whole tumor" (WT), which is defined by the union of the TC and the infiltrated tissue, which is the whole area that would be treated with radiation.

This first the data of 231 patient cases from 16 sites, and the resulting model was validated against the local data at each site. The second stage, called the *preliminary consensus model*, used the public initial model and incorporated more data from 2,471 patient cases from 35 sites, which improved its accuracy. The final stage, or *final consensus model*, used the updated model, and incorporated the largest amount of data from 6,314

patient cases (3,914,680 images) at 71 sites, across 6 continents, to further optimize and test for generalizability to unseen data.

As a control for each step, researchers excluded 20 percent of the total cases contributed by each participating site from the model training process and used as "local validation data." This allowed them to gauge the accuracy of the collaborative method. To further evaluate the generalizability of the models, six sites were not involved in any of the training stages to represent a completely unseen "out-of-sample" data population of 590 cases. Notably, the site at the American College of Radiology validated their model using data from a national clinical trial study.

Following model training the final consensus model garnered significant performance improvements against the collaborators' local validation [data](#). The final consensus model had an improvement of 27% in detecting ET boundaries, 33% in detecting TC boundaries, and 16% for WT boundary detection. The improved result is a clear indication of the benefit that can be afforded through access to more cases, not only to improve the model, but also to validate it.

Looking ahead, the authors hope that due to the generic methodology of federated learning, its applications in medical research can be far-reaching, applying not only to other cancers, but other conditions, like neurodegeneration, and beyond. They also anticipate more research to demonstrate that federated learning can abide by security and privacy protocols around the world.

More information: Federated learning enables big data for rare cancer boundary detection, *Nature Communications* (2022). [DOI: 10.1038/s41467-022-33407-5](https://doi.org/10.1038/s41467-022-33407-5)

Provided by Perelman School of Medicine at the University of Pennsylvania

Citation: Federated machine learning enables the largest brain tumor study to-date, without sharing patient data (2022, December 5) retrieved 30 January 2023 from <https://medicalxpress.com/news/2022-12-federated-machine-enables-largest-brain.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.