

ACCELERATING THE PATH TO PRECISION MEDICINE



ACCORDING to the U.S. Centers for Disease Control and Prevention, traumatic brain injury (TBI) affects millions of Americans each year, resulting in hundreds of thousands of emergency room visits and hospitalizations every year and as many as 150 deaths daily. Despite the ubiquity of TBI, targeted treatment, and, in many cases, a prognosis, for many patients, remains elusive. Clinically categorized as “mild, moderate, or severe,” the majority of TBIs, including concussions, fall into the mild category. However, people with mild TBI can later experience debilitating symptoms. Clinicians struggle to explain why someone experiences certain symptoms, but not others, and if they will ever go away. “Outcomes can range from a complete recovery after a devastating injury to debilitating depression or memory loss because of what, at first, seemed like an inconsequential incident,” says Shankar Sundaram, director of Lawrence Livermore’s Bioengineering Center and principal investigator of the Laboratory’s TBI research efforts.

With 100 billion neurons and 10,000 connections at the end of each neuron, the brain’s complexity confounds researchers. Yet, data analysis of the complex networks of neurons is mathematically well suited for applying Lawrence Livermore National Laboratory’s high-performance computing (HPC), machine learning, and data science expertise. In March 2018,

Diffusion magnetic resonance imaging (dMRI) measures the directional diffusion of water molecules in tissue, enabling the mapping of the underlying white matter structure of a living human brain. (Images courtesy of Tianhao Lanya Cai, University of California at San Francisco.)

Lawrence Livermore joined the Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) consortium. The national, multiyear, multidisciplinary effort, led by the University of California at San Francisco (UCSF) in collaboration with Lawrence Berkeley and Argonne national laboratories and other leading research organizations and universities, combines neuroimaging, blood-based biomarkers, clinical, and outcome data for thousands of TBI patients and advances supercomputing and machine learning with the aim of developing precision medicine diagnoses and treatments for TBI patients. So far, the results have accelerated, democratized, and simplified the data-processing workflow of TBI patient data and revealed the multifaceted nature of TBI. “Working with the TBI data set has presented us with one of the most complicated computing challenges we’ve ever faced,” says Sundaram. “We have been pushed, but I can say we’ve met the first set of challenges with success.”

Connectome Container

One of the first obstacles the Livermore team tackled was building a suitable data infrastructure to transfer, store, and provide access to huge quantities of complex, intricate multimodal patient data at high speeds to clinicians and TBI researchers. In the past, processing data from magnetic resonance imaging (MRI) of a single brain would take 24 hours before it could be read by a radiologist. Large-scale diffusion of MRI tractography—the visually stunning 3D-modeling technique depicting the brain’s nerve tracts—renders connectomes, comprehensive maps of neural connections or

disruptions in the brain. Connectomes present researchers with significant computational hurdles such as orchestrating multiple software packages with multiple dependencies and high computational costs. The prospect of sorting through and ingesting anonymized data from 3,000 TBI patients demanded the creation of a fast, streamlined system that could share data across laboratories participating in the TRACK-TBI consortium.

To solve this bottleneck and speed up data processing, Livermore software developer Joseph Moon and Peer-Timo Bremer, research scientist in the Laboratory’s Center for Applied Scientific Computing, developed the Massively Parallel, Portable, and Reproducible Tractography (MaPPeRTrac) workflow for HPC. MaPPeRTrac incorporates novel technologies to simplify and accelerate neuroimaging research. “One of the biggest obstacles to applying the latest data science techniques—graph analysis, statistical methods, and machine learning—is the complexity of neuroscience software,” says

Moon. “Dozens of tools, all designed by different teams around the world, must be connected to generate biophysically meaningful data.” MaPPeRTrac is a containerized parallel-compute infrastructure, which means it can be executed anywhere on top of any computing system without having to use multiple software packages or a supercomputer.

MaPPeRTrac fully automates the tractography workflow from management of raw MRI data to edge-density visualization of a connectome with the aim of providing clinicians in diverse hospital settings off-the-shelf access to clinically actionable data. The MaPPeRTrac container enables high-performance, parallel, parametrized, and portable generation of connectomes that is fast, efficient, well-tested, robust, and easy-to-use for the research community.

To lower the barrier to entry for users, Bremer’s team has also made MaPPeRTrac available on the open-source GitHub cloud so that any researcher—not only computer scientists—can generate connectomes using state-of-the-art software

libraries and HPC. “Users don’t have to be at a Department of Energy national lab with a high-performance computer to use MaPPeRTrac,” says Geoff Manley, director of the TRACK-TBI consortium at UCSF. “MaPPeRTrac democratizes the quality of TBI care clinicians can provide. Clinicians at a rural hospital, for example, can potentially access the fastest MRI analysis and predictive models. The team at Lawrence Livermore has vastly accelerated our data pipeline. Our collaboration and outcomes have been truly remarkable.”

Messy Data Challenge

The other major challenge was developing new algorithms that can process multimodal data—images, clinical assessments, biomarkers, and other data points—and produce reliable, interpretable diagnostics and prognostics even when some of the data is missing or uncertain. “An important aspect of applying machine learning in healthcare is managing missing data,” says Alan Kaplan, a Livermore research scientist working on the TBI project. “Many algorithms cannot utilize arbitrary ‘missing-ness’ patterns or require ad hoc preprocessing on ‘messy’ data. A research subject with incomplete data would typically be dropped from an algorithm’s analysis. But we all know that in real-world situations, life is messy, someone forgets to ask a question or leaves an answer blank. We had to figure out how to ensure that we don’t lose a subject if there are gaps in the data.”

The Livermore team created a statistical model, not just to impute but to account for “missing-ness” as well as random or deliberate omissions, considering if groupings or even sub-groupings might emerge despite gaps in the data. So far, the sub-groupings the algorithm has identified can predict outcomes using biomarkers, stratify patients into similar groups, stratify variables into similar groups, simulate patient data, and rank dependence of biomarkers on patient outcomes. The model allows for movement toward the idea of enhanced specificity among subgroups of TBI patients. “We now know that ‘mild, moderate, or severe’ is inadequate in terms of categorizing TBI,” says Kaplan. “The model we’ve developed incorporates data that portrays TBI’s truly multifaceted nature. The initial findings indicate a 20 percent improvement over baseline predictions. If validated, our work could help clinicians make more accurate prognoses and aid in decision making for early clinical management and determining targeted treatment interventions.”

Outcome groups	Patient clusters			
	1	2	3	4
Anxiety and depression	Green	Green	Red	Orange
Memory	Green	Red	Orange	Red
Global function	Green	Green	Orange	Red
Life satisfaction	Green	Orange	Red	Orange
Processing speed	Green	Orange	Orange	Red

The algorithm developed by Livermore data scientists can predict traumatic brain injury outcomes using biomarkers and rank dependence of these biomarkers from absence/low burden of specific symptoms (green) to some evidence/medium burden (orange), and presence/high burden (red), demonstrating the ability to provide enhanced clinical specificity beyond mild, moderate, or severe traumatic brain injuries.

Precision Medicine Goal

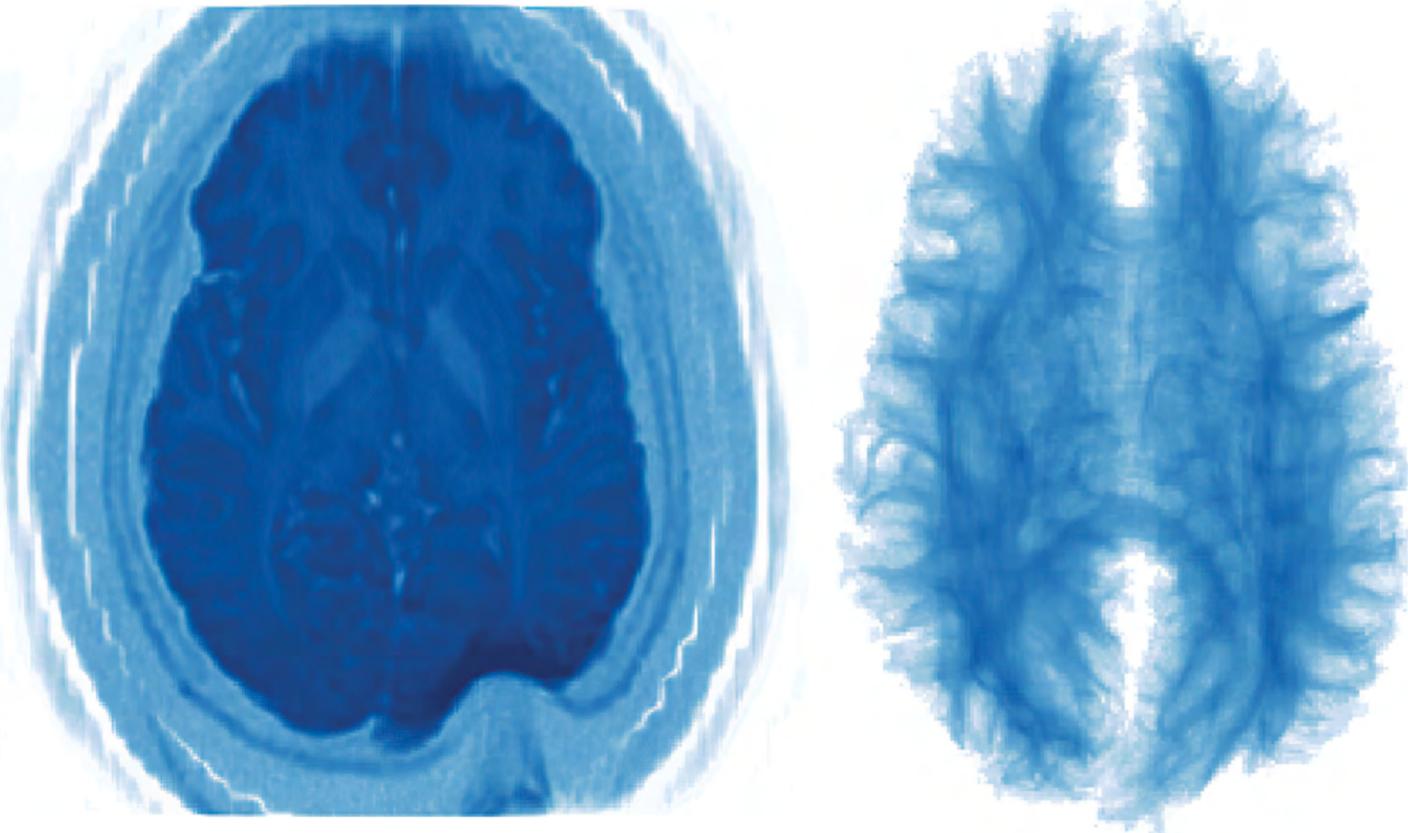
Having developed MaPPeRTrac to synthesize and process MRI connectome data as well as a machine-learning model to identify and organize swaths of TBI patient data despite missing data points, the Livermore team, working with TRACK-TBI, has a proof-of-concept. The team hopes to scale up the project by collaborating with other institutions including the National Collegiate Athletic Association, Veterans Affairs, and the Department of Defense to leverage their data sets of TBI patients and further engage the TBI research community. “This research isn’t about telling people there’s a cure, per se,” says Bremer. “But so often people experience overwhelming symptoms due to TBI, and clinicians have no way of predicting what the short- and long-term consequences may be. We’re hoping that our research can offer clinicians and their patients some insight into the cause of symptoms and the trajectory for treatment.”

“What’s really exciting about this work is that it could be used down the road to improve patient outcomes and have positive real-world impacts,” says Kaplan. Sundaram adds, “The tools are ready. With additional support, we will apply them to larger aggregated data sets to assess our predictive analytics and validate those biomarkers.”

—Genevieve Sexton

Key Words: brain tractography, connectome, high-performance computing (HPC), Massively Parallel, Portable, and Reproducible Tractography (MaPPeRTrac), magnetic resonance imaging (MRI), Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI), traumatic brain injury.

For further information contact Shankar Sundaram (925) 423-6468 (sundaram1@llnl.gov).



These images show (left) an MRI scan of a patient with traumatic brain injury and (right) a visualization of the simulated neural streamlines composing a brain’s connectome edge density. The latter was generated with the Laboratory’s Massively Parallel, Portable, and Reproducible Tractography workflow.