CASE STUDY

Health and Life Sciences
Data-Driven

Using Artificial Intelligence Solutions to Improve Patient Care

The University of California, San Francisco targets data-fueled insights for clinical medicine with Intel® Xeon® processors and BigDL* on Apache Spark*, an optimized deep learning framework for Intel® architecture

At a Glance:

The University of California, San Francisco (UCSF) is using Intel® Xeon® Scalable family processors and BigDL* on Apache Spark* to advance artificial intelligence (AI) for clinical medicine. These technologies allow UCSF to:

- Develop and train an AI model to identify a key precursor of osteoarthritis.
- Broaden its researchers' access to AI by deploying scalable, cost-effective infrastructure based on industry-standard technologies.
- Enhance its researchers' ability to solve clinical care challenges with a unified platform for AI and other analytics.

UCSF
University of California
San Francisco

"The deep learning and other big data analytics work we are doing with Intel moves us toward a new era of precision medicine that empowers clinicians to better predict outcomes, make more evidence-based treatment decisions, and transform the way we care for our patients. The potential is unparalleled."

—Rachael A Callcut, MD, MSPH, FACS
Director of Data Science, CDHI
Associate Professor of Surgery, UCSF
AI for Clinical Care

Tear the cartilage in your knee, and you could eventually face an impaired quality of life and major surgery because of osteoarthritis (OA). Now, researchers at the University of California, San Francisco (UCSF) Center for Digital Health Innovation (CDHI) are using deep learning, Intel® Xeon® processors, and BigDL* on Apache Spark* to develop an approach that may one day help patients avoid those consequences.

Challenge

As a leading academic medical center and research institution, the UCSF CDHI is improving clinical care through innovation in artificial intelligence (AI). To advance that work, university researchers require tools and cost-effective infrastructure to create new algorithms, train AI models, and run a variety of workloads. For the OA use case, UCSF researchers needed to develop and train a deep learning model that could examine sophisticated 3D medical images and identify those indicating torn knee cartilage.

Solution

The UCSF CDHI used BigDL on Apache Spark to develop algorithms and train models, and worked with Intel, Dell, and Cloudera to deploy a data analytics cluster based on the Intel® Xeon® Scalable processor family. Taking advantage of Intel data center technologies, UCSF broadens its researchers' ability to develop algorithms and models as well as to import and optimize models from other AI frameworks.

Results

The OA team is meeting its first-phase accuracy goals, demonstrating progress on a clinically significant problem with long-term potential to improve diagnosis, treatment planning, and clinician productivity. Intel's high-performance platform and tools are helping UCSF researchers speed time-to-results and make high-performance, AI capabilities broadly available to researchers.

AI for Clinical Care

OA is a serious degenerative condition and the most common joint disorder in the United States.[1] It compromises the quality of life for more than 30 million American adults and adds an estimated USD 62 billion annually in medical costs—and it's on the rise.[2]

A tear to the knee cartilage (the meniscus) can result from trauma. It can also be a sign of OA, or can exacerbate OA, and require eventual surgery. Clinicians diagnose a torn meniscus by reviewing and interpreting hundreds of high-resolution 3D magnetic resonance imaging (MRI) slices showing a patient's knee from different angles. Radiologists then assign a numerical score to indicate the presence and severity of a tear.

Solution Ingredients

- **Intel® Xeon® processors:** Intel's powerful, energy-efficient processors for data center and cloud computing environments
- **Cloudera Distribution of Apache Hadoop*:** Cloudera's open-source distribution of Apache Hadoop, enabling enterprises to store, process, and analyze multiple types of analytic workloads at large scale on industry-standard hardware
- **BigDL* on Apache Spark*:** An open-source, distributed library for creating deep learning applications that run on Apache Spark or Apache Hadoop clusters and deliver outstanding performance on Intel® technologies
- **Intel® Math Kernel Library (Intel® MKL):** A widely used library that accelerates math processing routines, increases application performance, and reduces development time for Intel® processor-based systems
- **TensorFlow*:** Open source deep learning and machine learning framework
- **TensorBoard*:** A suite of pen source visualization tools for debugging and optimizing TensorFlow programs

MRI and other 3D medical imaging studies are labor-intensive and time-consuming to interpret. They require levels of skill that are not always available, contributing to a global shortage of radiologists and often leaving patients and primary care teams waiting for results. In addition, radiologists interpret the images subjectively, without individualized high-resolution scoring. The resulting variability can affect treatment planning, follow-up assessment, and future research studies.

Combining UCSF's leadership in clinical medicine and data analytics, a UCSF CDHI research team is creating an AI solution to examine MRI results and objectively classify meniscus tears. The project is led by investigators Sharmila Majumdar, PhD, professor and vice chair of research in the UCSF Department of Radiology and Biomedical Imaging and professor in the Departments of Bioengineering and Therapeutic Sciences; and Valentina Pedoia, PhD, assistant professor in the UCSF Musculoskeletal and Imaging Research Group at the UCSF Department of Radiology and Biomedical Imaging. The project is one of dozens in which UCSF CDHI researchers are exploring how AI innovation can help solve critical challenges in clinical medicine.

The meniscus tear project is focused on developing an accurate, data-driven grading system of meniscus lesions. "This is a challenging and important problem that will have a number of benefits," said Majumdar. "It can provide results to patients immediately after scanning, help clinicians identify initiatory signs of osteoarthritis, and eliminate the variability that is present when multiple radiologists grade the same patient over time."

The meniscus research team is using Intel® technologies and BigDL on Apache Spark to create deep learning algorithms with a 3D convolutional neural network, and using existing MRI images to train the model to recognize meniscus tears. Once the model is sufficiently accurate, it can be integrated with other software and deployed as an inferencing solution to classify MRI studies in clinical or research environments.

The team is taking an incremental approach to model training, setting an initial goal of having the model determine whether the cartilage is normal or torn and achieving accuracy that matches or outperforms radiologists' analyses. "Even this binary classification has clinical relevance and importance," said Pedoia. "It will allow radiologists to quickly filter through cases and focus on patients who are at higher risk of meniscal tears or show early signs of osteoarthritis. It can also eliminate subjectivity issues, because the algorithm should make consistent decisions across data sets and studies."

Distributed Model Training for Cost-Effective, High-Performance AI

To support its ambitious AI strategies, the Center for Digital Health Innovation at UCSF deployed a powerful 10-node cluster that lets research teams develop algorithms, train models, and run other analytic workloads in a homogeneous, scalable, and CPU-based environment. The platform is based on the Intel® Xeon® processor E5 product family, CentOS Linux*, and the Cloudera Distribution of Apache Hadoop* (CDH*). The platform provides the flexibility to support multiple deep learning frameworks, the performance and capacity to run a variety of AI and big data workloads, and the scalability to handle future use case requirements. UCSF and Intel have teamed up in a broad collaboration to improve the delivery of frontline care through advances in deep learning analytics.

UCSF's meniscus classification team adopted BigDL on Apache Spark, a distributed, open source framework designed from the ground up for the Hadoop and Spark environments. BigDL uses the Intel® Math Kernel Library (Intel® MKL) to deliver outstanding performance on Intel® technology-based platforms. Data scientists and researchers can speed and simplify software development by writing their deep learning applications as standard Spark programs. They can develop algorithms, train models, and deploy their solutions on Spark or Hadoop clusters built on familiar Intel technologies, reducing both costs and complexity.

Choosing BigDL: High Performance and a Rich Feature Set

The UCSF CDHI chose BigDL on Apache Spark for its high throughput and extensive feature set. "We have a growing number of research teams working in AI, so scalability is important for us," said Rachael Callcut, MD, associate professor of surgery at UCSF and director of data science at the UCSF CDHI. "We need an affordable CPU-based development platform that allows us to maintain a homogeneous environment and has the performance and flexibility for

deep learning model training and inferencing. BigDL is on par with the other open source frameworks in its feature set, and it gives us a powerful development environment."

Figure 1 illustrates the scalable performance of BigDL and Intel technologies. Running the meniscus team's image recognition algorithms and model-training data set on BigDL on Apache Spark v0.3, the solution scaled along two critical dimensions. It enabled faster training as testers allocated more compute cores, as well as when they proportionately increased the batch size, i.e., the number of examples in the training data set (Figure 2). This ability to use resources efficiently can help reduce the time needed to train deep learning models.

The model ran for 20 training cycles or epochs. Performance tests conducted by Intel on the UCSF system also showed BigDL providing improvements to throughput in processing ImageNet* images than on CaffeOnSpark*, effectively reducing training time.

Intel has been working closely with the open source community to add requested features to BigDL, including capabilities that save developers time in developing 3D convolutional neural network models and make it easier to preprocess 3D images coming from diverse sources. Developers can also use BigDL to import models developed in other frameworks and take advantage of tools and applications such as TensorBoard* and Jupyter Notebook*. UCSF researchers are using TensorBoard to monitor their training and validation metrics and have integrated algorithms they had previously developed with TensorFlow*.

Simplify Deployment of BigDL on Apache Spark with Intel® Select Solutions

Intel offers BigDL on Apache Spark as part of the Intel® Select Solutions program. Intel Select Solutions are verified hardware and software stacks optimized across compute, storage, and networking resources for specific workloads. Built on Intel Xeon Scalable processors, Intel Select Solutions help ensure enterprises get the performance, agility, and security they require. These solutions are verified to meet a specified minimum level of workload-optimized performance capabilities, helping reduce the time it will take to evaluate, select, and deploy the Intel Select Solution for BigDL on Apache Spark.

About Intel® Select Solutions

Intel® Select Solutions are verified hardware and software stacks that are optimized for specific software workloads across compute, storage, and network. These solutions are developed from deep Intel experience with industry solution providers, in addition to extensive collaboration with the world's leading data center and service providers.

To qualify as an Intel Select Solution, solution providers must:

1. Follow the software and hardware stack requirements outlined by Intel.
2. Replicate or exceed Intel's minimum benchmark-performance thresholds.
3. Provide resources, services, or documentation to facilitate customer deployment.

Solution providers can develop their own optimizations to add further value to their solutions.

Scalable Performance with BigDL* and an Intel® Xeon® Processor-Based System

Figure 1. BigDL* on Apache Spark* and Intel® technologies accelerated training as testers allocated more compute cores.

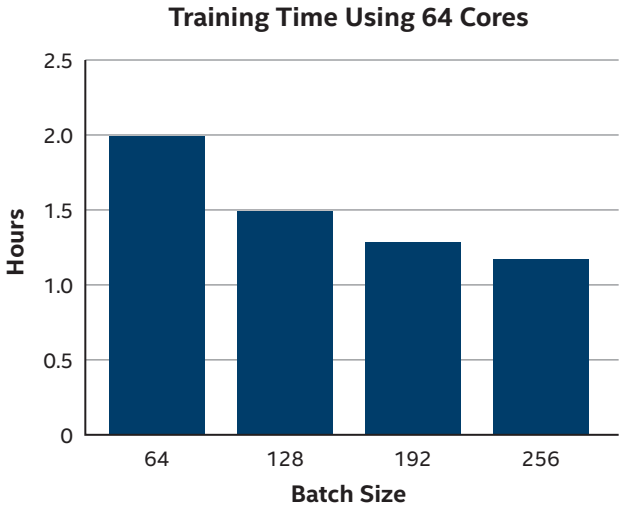

Figure 2. Increased the number of examples in the training data set proportionally.

Conclusion: Imaging for Precision Medicine

A leader in academic and clinical medicine, UCSF is advancing the use of AI to improve medical imaging and patient care. By using BigDL and Intel technologies, the UCSF CDHI provides researchers with a high-performance environment for deep learning and other data-intensive analytics workloads. The institution stretches its research dollars and makes it feasible for a broad range of researchers to explore how they can incorporate deep learning and other forms of AI into their disciplines. Using this environment, CDHI researchers are making progress on significant clinical research challenges, including work that may one day increase clinical efficiency and help patients avoid the pain—and health systems reduce the costs—of OA.

"The deep learning and other big data analytics work we are doing with Intel is helping us start to turn a wealth of clinical and other data into actionable insights," says Callcut. "It moves us toward a new era of precision medicine that empowers clinicians to better predict outcomes, make more evidence-based treatment decisions, and transform the way we care for our patients. The potential is unparalleled."

Find the solution that is right for your organization. Contact your Intel representative or visit **intel.com/selectsolutions**.

Spotlight on the UCSF

University of California, San Francisco (UCSF) is a leading university dedicated to promoting health worldwide through advanced biomedical research, graduate-level education in the life sciences and health professions, and excellence in patient care. It includes top-ranked graduate schools of dentistry, medicine, nursing, and pharmacy; a graduate division with nationally renowned programs in basic, biomedical, translational and population sciences; a preeminent biomedical research enterprise; and a network of hospitals and affiliated healthcare providers throughout the San Francisco Bay Area. The university has been the top public recipient of biomedical research grants from the National Institutes of Health (NIH) for the sixth consecutive year in 2016, and the second-highest recipient among all public and private institutions nationwide, according to annual figures from the NIH.[3] For more information about UCSF and the Center for Digital Health Innovation, visit ucsf.edu and centerfordigitalhealthinnovation.org.

Learn More

Solution Brief

- Inferencing Solution Simplifies AI Adoption for Medical Imaging

Relevant Intel Products and Solutions

- Intel® Xeon® Scalable processors
- Intel® MKL
- Intel® Select Solutions

Other Technologies

- BigDL on Apache Spark*
- Cloudera Distribution of Apache Hadoop*
- Apache Spark*
- TensorFlow*
- TensorBoard*

[1] Felson DT, Lawrence RC, Dieppe PA, et al. Osteoarthritis: New Insights. Part 1: The Disease and Its Risk Factors. Ann Intern Med. 2000;133(8):635–46. ncbi.nlm.nih.gov/pubmed/11033593

[2] Osteoarthritis Action Alliance, Cost of Osteoarthritis, April 5 2017, oaaction.unc.edu/policy-solutions/cost-of-osteoarthritis

[3] Nicholas Weiler, UCSF News Center, UCSF Is Top Public Recipient of NIH Research Funding for Sixth Consecutive Year, March 27, 2017. ucsf.edu/news/2017/03/406096/ucsf-top-public-recipient-nih-research-funding-6th-consecutive-year

1018/JSTAR/KC/PDF 336720-001US